

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 34.1497 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-1
Perfect score: 1184
Sequence: 1 GPGEPGPTGLPQPPGERGG.....GEQVPGDLGAPGSPGAG 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1184	100.0	209	ADM48390	Adm48390 Human rec
2	1178	99.5	617	ADM48391	Adm48391 Human rec
3	1178	99.5	617	ADQ26217	Adq26217 Human gel
4	1178	99.5	821	ADM48392	Adm48392 Human rec
5	1149	97.0	501	AAB68057	Aab68057 Amino aci
6	1149	97.0	501	AAB68057	Aab68057 Amino aci
7	1149	97.0	501	ADQ26217	Adq26217 Human gel
8	1149	97.0	1057	ADQ26217	Adq26217 Human gel
9	1149	97.0	1057	ADQ26217	Adq26217 Human gel
10	1149	97.0	1058	ADQ26217	Adq26217 Human gel
11	1149	97.0	1107	ADQ26217	Adq26217 Human gel
12	1149	97.0	1107	ADQ26217	Adq26217 Human gel
13	1149	97.0	1161	ADQ26217	Adq26217 Human gel
14	1149	97.0	1169	ADQ26217	Adq26217 Human gel
15	1149	97.0	1169	ADQ26217	Adq26217 Human gel
16	1149	97.0	1171	ADQ26217	Adq26217 Human gel
17	1149	97.0	1171	ADQ26217	Adq26217 Human gel
18	1149	97.0	1211	ADQ26217	Adq26217 Human gel
19	1149	97.0	1226	ADQ26217	Adq26217 Human gel
20	1149	97.0	1341	ADQ26217	Adq26217 Human gel
21	1149	97.0	1341	ADQ26217	Adq26217 Human gel
22	1149	97.0	1341	ADQ26217	Adq26217 Human gel
23	1149	97.0	1341	ADQ26217	Adq26217 Human gel
24	1149	97.0	1341	ADQ26217	Adq26217 Human gel
25	1149	97.0	1341	ADQ26217	Adq26217 Human gel

26	1149	97.0	1388	2	AAR89471	Aar89471 Collagen/
27	1149	97.0	1411	3	AAY56800	Aay56800 Human pre
28	1149	97.0	1451	5	ABG3947	Abg3947 Human pol
29	1149	97.0	1464	2	AAW68485	Aaw68485 Human rec
30	1149	97.0	1464	4	AAB82454	Aab82454 Human pro
31	1149	97.0	1464	31	AAU14136	Aau14136 Human nov
32	1149	97.0	1464	5	ABB90764	Abb90764 Human tum
33	1149	97.0	1464	5	ABP68610	Abp68610 Human pan
34	1149	97.0	1464	6	ABU54471	Abu54471 Human tum
35	1149	97.0	1464	6	ABR47417	AbR47417 Breast ca
36	1149	97.0	1464	6	ABR32064	AbR32064 Human cer
37	1149	97.0	1464	7	ADD14142	Add14142 Human src
38	1149	97.0	1464	7	ADD45059	Add45059 Human pro
39	1149	97.0	1464	7	ADD45055	Add45055 Human pro
40	1149	97.0	1464	7	ADD45051	Add45051 Human pro
41	1149	97.0	1464	7	ADP65246	Adp65246 Human alp
42	1149	97.0	1464	7	ADQ19470	Adq19470 Human sof
43	1149	97.0	1464	8	ADQ29653	Adq29653 Human col
44	1149	97.0	1464	8	ADQ29653	Adq29653 Human col
45	1149	97.0	1464	8	ADR16800	Adr16800 Human col

ALIGNMENTS

RESULT 1
ADM48390
ID ADM48390 standard; protein; 209 AA.

XX AC ADM48390;
XX DT 03-JUN-2004 (first entry)
XX DE Human recombinant gelatin-like polypeptide Hu-1.
XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.
XX OS Homo sapiens.
XX FN EPI398324-A1.
XX PD 17-MAR-2004.
XX PF 11-SEP-2002; 2002EP-00078745.
XX PR 11-SEP-2002; 2002EP-00078745.
XX PA (FUJF) FUJI PHOTO FILM BV.
XX FI Bouwstra JB, Toda Y;
XX DR WPI; 2004-229415/22.
XX PT Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.
XX PS Example 1; SEQ ID NO 1; 31pp; English.

The present sequence is the protein sequence of human recombinant gelatin-like protein Hu-1. This has a molecular weight of 18.4 kDa and an isoelectric point of 5.35. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgE antibodies. The gelatin-like proteins can be covalently attached to

CC pharmaceutically active compounds. After administration, the coupled
CC medicament will not diffuse from the circulating blood into the
CC interstitium. Clearance by liver and kidney will be kept to a minimum,
CC ensuring a more constant plasma level of the medicament. Suitable
CC medicaments include those involved in intervening blood clotting,
CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,
CC immune responses, and blood levels of messenger molecules such as
CC hormones.
XX
SQ Sequence 209 AA;

Query Match 100.0%; Score 1184; DB 8; Length 209;
Best Local Similarity 100.0%; Pred. No. 9.5e-70;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPGRGPGSRGPFADGVAGPKGAGERSGPGAGKSPGAGRPP 60
DB 1 GPPGEPGPTGLPGRGPGSRGPFADGVAGPKGAGERSGPGAGKSPGAGRPP 60

QY 61 GEAGLPKAKGLTSGSPGSDGKTGPPGAGQDGRFPFGPGARQAGVMGFPQPKGAA 120
DB 61 GEAGLPKAKGLTSGSPGSDGKTGPPGAGQDGRFPFGPGARQAGVMGFPQPKGAA 120

QY 121 GEPKAGRGVPPGAVGAGPKAGDGEAGQAGQPPGAGPAGERGEGQAGSPQFQGLPGPA 180
DB 121 GEPKAGRGVPPGAVGAGPKAGDGEAGQAGQPPGAGPAGERGEGQAGSPQFQGLPGPA 180

QY 181 GPPGAGKPGEGQGVGDLDGAPGSPGAGG 209
DB 181 GPPGAGKPGEGQGVGDLDGAPGSPGAGG 209

RESULT 2
ADM48391
ID ADM48391 standard; protein; 617 AA.
XX
AC ADM48391;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human recombinant gelatin-like polypeptide Hu-3.
XX
KW Plasma substitute; Gelatin-like protein; plasma expander; human.
XX
OS Homo sapiens.
XX
PN EP1398324-A1.
XX
PD 17-MAR-2004.
XX
PF 11-SEP-2002; 2002EP-00078745.
XX
PR 11-SEP-2002; 2002EP-00078745.
XX
PA (FUJF) FUJI PHOTO FILM BV.
XX
PI Bouwstra JB, Toda Y;
XX
DR WPI; 2004-229415/22.
XX
PT Composition useful as substitute for plasma, comprises solution of saline
XX and recombinant gelatin-like protein having colloid osmotic function.
XX
PS Example 1; SEQ ID NO 2; 31pp; English.
XX
CC The present sequence is the protein sequence of recombinant gelatin-like
CC protein Hu-3. This is a trimer of human recombinant gelatin-like protein
CC Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma
CC comprises a solution of saline and a protein having a colloid osmotic
CC function. The protein is a recombinant gelatin-like protein with a
CC molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric
CC point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The
CC recombinant gelatin-like protein, or a dimer, trimer or tetramer of the

CC protein, is useful as a plasma expander that has a lower clearance rate
CC from blood circulation, provides better and predictable regulation of
CC clearance rate and which is less susceptible to proteolytic degradation
CC than presently used gelatin derivatives. Recombinant gelatin-like
CC proteins that are in essence free of hydroxyproline do not give rise to
CC an immunological reaction with blood samples containing IgE antibodies.
CC The gelatin-like proteins can be covalently attached to pharmaceutical
CC active compounds. After administration, the coupled medicament will not
CC diffuse from the circulating blood into the interstitium. Clearance by
CC liver and kidney will be kept to a minimum, ensuring a more constant
CC plasma level of the medicament. Suitable medicaments include those
CC involved in intervening blood clotting, vasodilation, function of
CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood
CC levels of messenger molecules such as hormones.
XX
SQ Sequence 617 AA;

Query Match 99.5%; Score 1178; DB 8; Length 617;
Best Local Similarity 100.0%; Pred. No. 5.7e-69;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPGRGPGSRGPFADGVAGPKGAGERSGPGAGKSPGAGRPP 60
DB 1 GPPGEPGPTGLPGRGPGSRGPFADGVAGPKGAGERSGPGAGKSPGAGRPP 60

QY 61 GEAGLPKAKGLTSGSPGSDGKTGPPGAGQDGRFPFGPGARQAGVMGFPQPKGAA 120
DB 61 GEAGLPKAKGLTSGSPGSDGKTGPPGAGQDGRFPFGPGARQAGVMGFPQPKGAA 120

QY 121 GEPKAGRGVPPGAVGAGPKAGDGEAGQAGQPPGAGPAGERGEGQAGSPQFQGLPGPA 180
DB 121 GEPKAGRGVPPGAVGAGPKAGDGEAGQAGQPPGAGPAGERGEGQAGSPQFQGLPGPA 180

QY 181 GPPGAGKPGEGQGVGDLDGAPGSPGAG 208
DB 181 GPPGAGKPGEGQGVGDLDGAPGSPGAG 208

RESULT 3
ADQ26217
ID ADQ26217 standard; protein; 617 AA.
XX
AC ADQ26217;
XX
DT 23-SEP-2004 (first entry)
XX
DE Human gelatine-like polypeptide Hu-3.
XX
KW Human; gelatine-like protein; Hu-3; microcarrier; cell culture.
XX
OS Homo sapiens.
XX
PN WO2004056976-A2.
XX
PD 08-JUL-2004.
XX
PF 23-DEC-2003; 2003WO-NL000922.
XX
PR 23-DEC-2002; 2002EP-00080539.
XX
PA (FUJF) FUJI PHOTO FILM BV.
XX
PI Bouwstra JB, Van Es AJJ, Toda Y;
XX
DR WPI; 2004-507711/48.
XX
PT Preparing cell culture support useful for culturing anchorage dependent
XX cells, involves coating microcarrier bead with gelatine or gelatine-like
XX protein having specific molecular weight.
XX
PS Example 1; SEQ ID NO 1; 19pp; English.
XX
CC The present sequence is the protein sequence of human recombinant gelatin

XX

XX

PD 17-MAY-2001.
 PF 10-NOV-2000; 2000WO-US030843.
 XX
 PR 12-NOV-1999; 99US-0165114P.
 PR 15-MAY-2000; 2000US-0204437P.
 XX
 PA (FIBR-) FIBROGEN INC.
 XX
 XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 XX WPI; 2001-308784/32.
 DR
 XX Vaccine formulations (1) comprising recombinant human gelatin, useful for
 PT vaccinating against e.g. mumps, measles, rubella, tetanus, rabies and
 PT cholera, the gelatin is non-immunogenic and confers stability at ambient
 PT temperatures.
 XX
 XX Claim 11; Page 114-116; 130pp; English.
 XX
 CC The present sequence represents a human recombinant gelatin polypeptide.
 CC The recombinant gelatin polypeptide is used to produce vaccine
 CC formulations of the invention. The recombinant human gelatin is non-
 CC immunogenic (therefore reducing anaphylactic reactions) and confers
 CC stability at ambient temperatures. The vaccine formulation comprises a
 CC vaccine formulated for the prevention of a disease selected from vaccinia
 CC virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
 CC diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
 CC (whooping cough), Bacille Calmette-Guérin (BCG, tuberculosis),
 CC haemophilus influenzae meningitis, rabies, cholera, Japanese
 CC encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
 CC adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
 CC respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
 CC herpes virus (Marek's disease), influenza and/or anthrax
 XX
 SQ Sequence 501 AA;
 Query Match 97.0%; Score 1149; DB 4; Length 501;
 Best Local Similarity 93.6%; Pred. No. 3.7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTGLPGRGGRGSGRFGAGVAGKPGKPGAGRSGPGPA 48
 DB 280 GPPGAGEGKRGARGEPGTGLPGRGGRGSGRFGAGVAGKPGKPGAGRSGPGPA 339
 QY 49 GPKSGPGEAGRPGEGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPGARGQA 108
 DB 340 GPKSGPGEAGRPGEGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPGARGQA 399
 QY 109 GVMGFPKPGKGAAGEPGKAGRGVPGPPGAVGPKDGEAGAGQPPGAPGAGERGEOQPA 168
 DB 400 GVMGFPKPGKGAAGEPGKAGRGVPGPPGAVGPKDGEAGAGQPPGAPGAGERGEOQPA 459
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGPAG 208
 DB 460 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGARG 499
 RESULT 6
 AA020703
 ID AA020703 standard; protein; 501 AA.
 XX
 AC AA020703;
 XX
 XX 06-AUG-2001 (first entry)
 XX
 XX Human alpha (I) type I collagen helical domain (residues 179-679).
 XX
 KW Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
 KW encapsulant; film-forming agent; moisturising agent; thickening agent;
 KW gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
 KW plasma expander; colloidal volume replacement material; graft coating;
 KW medical sponge; medical plug; micro-carrier; edible composition;

KW protein supplement; fat substitute; nutritional supplement; cell culture;
 KW edible coating; cosmetic; vaccine; therapy; arthritis; achrosis;
 KW cartilage degeneration; joint flexibility; food industry; beverage;
 KW alpha (I) type I collagen.
 OS Homo sapiens.
 XX
 XX WO200134646-A2.
 PN
 XX 17-MAY-2001.
 PD
 XX 10-NOV-2000; 2000WO-US030791.
 PF
 XX 12-NOV-1999; 99US-0165114P.
 PR
 PR 15-MAY-2000; 2000US-0204437P.
 XX
 PA (FIBR-) FIBROGEN INC.
 XX
 XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 FI WPI; 2001-329072/34.
 XX
 DR Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
 XX prepared recombinantly.
 PT
 PT Claim 21; Page 121-123; 137pp; English.
 PS
 XX The patent discloses recombinant human gelatin which is useful in various
 CC compositions including binding agents, encapsulants, stabilising agents,
 CC film-forming agents, moisturising agents, emulsifiers, thickening agents,
 CC gelling agents, colloidal agents, adhesive agents, pharmaceutical
 CC compositions, hard gel capsules, soft gel capsules, plasma expander,
 CC colloidal volume replacement materials, graft coatings, medical sponges,
 CC medical plugs, pharmaceutical stabilisers, micro-carriers, edible
 CC compositions, protein supplements, fat substitutes, nutritional
 CC supplements, edible coatings, photographic compositions, cosmetic
 CC compositions, industrial composition, cell culture compositions and
 CC compositions for use in the laboratory. Pharmaceutical compositions
 CC comprising recombinant gelatin are used as vaccines. They are also used
 CC to treat various joint conditions such as arthritis, achrosis and other
 CC conditions related to the degeneration of cartilage and joint
 CC flexibility. Recombinant gelatin is also used in food and beverage
 CC industries. The present sequence is human alpha (I) type I collagen
 CC helical domain (residues 179-679). This sequence is a recombinant gelatin
 XX
 SQ Sequence 501 AA;
 Query Match 97.0%; Score 1149; DB 4; Length 501;
 Best Local Similarity 93.6%; Pred. No. 3.7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTGLPGRGGRGSGRFGAGVAGKPGKPGAGRSGPGPA 48
 DB 280 GPPGAGEGKRGARGEPGTGLPGRGGRGSGRFGAGVAGKPGKPGAGRSGPGPA 339
 QY 49 GPKSGPGEAGRPGEGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPGARGQA 108
 DB 340 GPKSGPGEAGRPGEGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPGARGQA 399
 QY 109 GVMGFPKPGKGAAGEPGKAGRGVPGPPGAVGPKDGEAGAGQPPGAPGAGERGEOQPA 168
 DB 400 GVMGFPKPGKGAAGEPGKAGRGVPGPPGAVGPKDGEAGAGQPPGAPGAGERGEOQPA 459
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGPAG 208
 DB 460 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGARG 499
 RESULT 7
 AD084290
 ID AD084290 standard; protein; 501 AA.
 XX
 AC AD084290;

XX DT 04-DEC-2003 (first entry)
 XX DE Recombinant gelatin #3.
 XX KW viricide; antitussive; tuberculostatic; antibacterial; antiinflammatory;
 KW hepatotropic; vaccine; recombinant gelatin; vaccinia virus; small pox;
 KW polio virus; Salk and sabin; mumps; measles; rubella; diphtheria;
 KW tetanus; varicella-zoster; chicken pox; shingles; pertussis;
 KW whooping cough; Bacille Calmette-Guerin; BCG; tuberculosis;
 KW Haemophilus influenzae; meningitis; rabies; cholera;
 KW Japanese encephalitis virus; Salmonella typhi; Shigella; hepatitis A;
 KW hepatitis B; adenovirus; yellow fever; foot and mouth disease;
 KW herpes simplex virus; respiratory syncytial virus; rotavirus; dengue;
 KW West Nile virus; turkey herpes virus; Marek's disease; influenza;
 KW anthrax; collagen type 1; alpha1(I).
 XX OS Synthetic.
 XX OS Homo sapiens.
 XX PN US2003064074-A1.
 XX PD 03-APR-2003.
 XX PF 30-AUG-2002; 2002US-00232175.
 XX PR 12-NOV-1999; 99US-0165114P.
 XX PR 15-MAY-2000; 2000US-0204437P.
 XX PR 10-NOV-2000; 2000US-00710249.
 XX PA (CHAN/) CHANG R C.
 XX PA (KIVI/) KIVIRIKKO K I.
 XX PA (NEFF/) NEFF T B.
 XX PA (OLSE/) OLSEN D R.
 XX PA (POLA/) POLAREK J W.
 XX PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 XX WPI; 2003-540775/51.
 XX PT New vaccine composition comprising a recombinant gelatin and an antigenic
 PT agent, useful for preventing e.g. polio virus, mumps, measles, rubella,
 PT diphtheria, tetanus, chicken pox/shingles, pertussis, cholera, rotavirus
 PT or dengue.
 XX PS Claim 16; Page 36-38; 63pp; English.
 XX CC The invention describes a vaccine composition comprising a recombinant
 CC gelatin, and an antigenic agent. The vaccine can be delivered by
 CC injection, through nasal, oral, transdermal or mucosal routes, or through
 CC deep lung delivery. Administration may also be oral, rectal,
 CC transcutaneous, intramedullary, intrathecal, intraventricular, intravenous,
 CC intraperitoneal, intranasal or intraocular injection). The vaccine
 CC composition is formulated for the prevention of vaccinia virus (small
 CC pox), polio virus (Salk and Sabin), mumps, measles, rubella, diphtheria,
 CC tetanus, Varicella-Zoster (chicken pox/shingles), pertussis (whooping
 CC cough), Bacille Calmette-Guerin (BCG, tuberculosis), Haemophilus
 CC influenzae meningitis, rabies, cholera, Japanese encephalitis virus,
 CC Salmonella typhi, Shigella hepatitis A, hepatitis B, adenovirus, yellow
 CC fever, foot and mouth disease, herpes simplex virus, respiratory
 CC syncytial virus, rotavirus, dengue, West Nile virus, Turkey herpes virus
 CC (Marek's disease), influenza, and anthrax. This is the amino acid
 CC sequence of a recombinant gelatin used in the creation of a vaccine
 CC composition of the invention.
 XX SQ Sequence 501 AA;
 XX Query Match 97.0%; Score 1149; DB 7; Length 501;
 XX Best Local Similarity 93.6%; Pred. No. 3.7e-67;
 XX Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPGPTGLPGPGERGPGSRGFGADGVAGKPGAGERGSGPQA 48

Db 280 GPGPAGBEGKRGARGEPGPTGLPGPGERGPGSRGFGADGVAGKPGAGERGSGPQA 339
 Qy 49 GPKSGPGEAGRPGEGAGLPGAKGLTGTSGSPGPPGKTPGPPAGQDGRPPGPPGARGQA 108
 Db 340 GPKSGPGEAGRPGEGAGLPGAKGLTGTSGSPGPPGKTPGPPAGQDGRPPGPPGARGQA 399
 Qy 109 GVMGFPKPKAAGEPGKAGRGVPGPCAVGPAGKDGAGAGQPPGPPGAGERGEGQA 168
 Db 400 GVMGFPKPKAAGEPGKAGRGVPGPCAVGPAGKDGAGAGQPPGPPGAGERGEGQA 459
 Qy 169 GSPGFQGLPGPAGPPGKAGRGVPGQGVPGDLGAPGSPGAG 208
 Db 460 GSPGFQGLPGPAGPPGKAGRGVPGQGVPGDLGAPGSPGAG 499
 RESULT 8
 AAY84541
 ID AAY84541 standard; protein; 1057 AA.
 XX AC AAY84541;
 XX DT 25-JUL-2000 (first entry)
 XX DE Amino acid sequence of a human collagen 1 (alpha1) protein.
 XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation.
 XX OS Homo sapiens.
 XX PN EP992586-A2.
 XX PD 12-APR-2000.
 XX PF 07-OCT-1999; 99EP-00119184.
 XX PR 09-OCT-1998; 98US-00169768.
 XX PA (USSU) US SURGICAL CORP.
 XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX WPI; 2000-259138/23.
 XX DR N-PSDB; AAA12502.
 XX PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX PS Disclosure; Fig 27A-E; 260pp; English.
 XX CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a human collagen 1 (alpha1) protein, which
 CC may be produced using the method of the invention

XX SQ Sequence 1057 AA;
 Query Match 97.0%; Score 1149; DB 3; Length 1057;
 Best Local Similarity 93.6%; Pred. No. 6.8e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GSPGPTGLPGPPGGRGSGRFPAGDGVAGPKGPGAGRSGSPGA 48
 DB 297 GPPGAGEGKRGARGEGPTGLPGPPGGRGSGRFPAGDGVAGPKGPGAGRSGSPGA 356
 QY 49 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGPAGAGDGRPPGPPGARGQA 108
 DB 357 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGPAGAGDGRPPGPPGARGQA 416
 QY 109 GVMGFPFGKGAAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 168
 DB 417 GVMGFPFGKGAAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 476
 QY 169 GSPGFQGLPGPAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 208
 DB 477 GSPGFQGLPGPAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 516
 RESULT 9
 AAY84544
 ID AAY84544 standard; protein; 1057 AA.
 AC AAY84544;
 DT 25-JUL-2000 (first entry)
 DE A human collagen 1 (alpha1) protein helical region.
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 collagen; fibrinogen; fibronectin; post translational hydroxylation.
 OS Homo sapiens.
 PN EP992586-A2.
 PD 12-APR-2000.
 PF 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 PA (USSU) US SURGICAL CORP.
 PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 DR WPI; 2000-259138/23.
 DR N-PSDB; AAA12503.
 XX Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX Example 10; Fig 39A-E; 260pp; English.
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-

CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents human collagen 1 (alpha1) helical region,
 CC which may be produced using the method of the invention
 XX Sequence 1057 AA;
 Query Match 97.0%; Score 1149; DB 3; Length 1057;
 Best Local Similarity 93.6%; Pred. No. 6.8e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GSPGPTGLPGPPGGRGSGRFPAGDGVAGPKGPGAGRSGSPGA 48
 DB 297 GPPGAGEGKRGARGEGPTGLPGPPGGRGSGRFPAGDGVAGPKGPGAGRSGSPGA 356
 QY 49 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGPAGAGDGRPPGPPGARGQA 108
 DB 357 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGPAGAGDGRPPGPPGARGQA 416
 QY 109 GVMGFPFGKGAAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 168
 DB 417 GVMGFPFGKGAAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 476
 QY 169 GSPGFQGLPGPAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 208
 DB 477 GSPGFQGLPGPAGPFGKAGRGVFPFGAVGPKGDEAGAGQPPGPPGAGRGGQGA 516
 RESULT 10
 AAY84403
 ID AAY84403 standard; protein; 1058 AA.
 AC AAY84403;
 DT 12-JUL-2000 (first entry)
 DE Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
 KW Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
 KW collagen; mussel adhesive protein; bioadhesive.
 OS Homo sapiens.
 PN WO200014201-A1.
 PD 16-MAR-2000.
 PF 07-SEP-1999; 99WO-US020462.
 PR 09-SEP-1998; 98US-0099652P.
 XX (USSU) US SURGICAL CORP.
 XX (PAOL/) PAOLELLA D N.
 XX (GRUS/) GRUSKIN E A.
 XX (BUEC/) BUECHTER D D.
 PI Paolella DN, Gruskin EA, Buechter DD;
 XX WPI; 2000-271051/23.
 DR N-PSDB; AAZ99843.
 XX Incorporating non-natural amino acid into polypeptide, useful e.g. for
 PT production of bioadhesives, by epoxidation or substitution of
 PT dehydroproline residues.
 XX Disclosure; Fig 6; 66pp; English.
 PS The present sequence represents a human type 1 (alpha1) collagen protein.
 CC Peptides derived from the protein were used to demonstrate incorporation

CC of 3,4-dehydro-L-proline into the peptide, using the method of the
 CC invention. The specification describes a method for the incorporation of
 CC non-natural amino acid into a polypeptide. The method comprises reacting
 CC at least one 3,4-dehydroproline residue in the polypeptide with an
 CC epoxidation reagent from a polypeptide containing at least one 3,4-
 CC epoxypoline residue. The method is used for studying the effects of non-
 CC natural amino acids on structure and function of polypeptides. The method
 CC is also useful for commercial production of collagen or mussel adhesive
 CC proteins (which are useful as bioadhesives), and for incorporating a wide
 CC variety of groups, including therapeutic ligands and biological probes,
 CC into polypeptides
 XX
 XX Sequence 1058 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1058;
 Best Local Similarity 93.6%; Pred. No. 6.8e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGPTGLPGRGSGRPGPGADGVAGPKPAGERGSPCPA 48
 DB 298 GPPGAGEGKRGAGEPGTGLPGRGSGRPGPGADGVAGPKPAGERGSPCPA 357
 QY 49 GPKGSPGAGRPGAGLPGAGLGTGSPGSGPGDGTGTPGAGQDGRPPGPPGARGQA 108
 DB 358 GPKGSPGAGRPGAGLPGAGLGTGSPGSGPGDGTGTPGAGQDGRPPGPPGARGQA 417
 QY 109 GVMGFPKGAAGSPGKAGRGVPGPGAVGAGKDGAGAGQPPGAGPAGERGEGCPA 168
 DB 418 GVMGFPKGAAGSPGKAGRGVPGPGAVGAGKDGAGAGQPPGAGPAGERGEGCPA 477
 QY 169 GSPGFQGLPGAGPGEAGKPGEGVPGDLGAPGSPGAG 208
 DB 478 GSPGFQGLPGAGPGEAGKPGEGVPGDLGAPGSPGARG 517

RESULT 11
 AAR89472
 ID AAR89472 standard; protein; 1107 AA.
 XX
 XX AAR89472;
 XX
 XX 01-OCT-1996 (first entry)
 XX
 XX Collagen/decorin(aa46-93) fusion protein.
 XX
 XX Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
 XX bone formation; tissue repair; fusion protein.
 XX Synthetic.
 XX
 XX Key Location/Qualifiers
 XX Domain 1..1057
 XX /label= Collagen-IA
 XX /note= "collagen IA alpha-helical domain"
 XX
 XX Misc-difference 887
 XX /note= "unidentified amino acid"
 XX
 XX Misc-difference 890
 XX /note= "unidentified amino acid"
 XX
 XX Peptide 1058..1059
 XX /label= linker_peptide
 XX
 XX Domain 1060..1107
 XX /label= Decorin
 XX /note= "amino acids P46 to G93 of mature decorin"
 XX
 XX CA2151547-A.
 XX
 XX 11-DEC-1995.
 XX
 XX 12-JUN-1995; 95CA-02151547.
 XX
 XX 10-JUN-1994; 94US-00259263.
 XX
 XX (USSU) US SURGICAL CORP.

XX Gruskin EA, Espino P;
 XX
 XX WPI; 1996-140144/15.
 DR N-PSDB; AAT16518.
 XX
 XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
 PT - and cellular regulatory factor domain, partic. useful as osteogenic
 XX agents, also related vectors, transformed cells and chimaeric proteins.
 XX Disclosure; Fig 8; 59pp; English.
 XX
 XX A fusion protein (AAR89472) comprises the alpha-helical region of human
 CC collagen I(a) linked to amino acids 46-93 of human mature dermatan
 CC sulphate proteoglycan (decorin). It can be expressed in Escherichia coli
 CC transformants carrying a vector incorporating a chimeric gene (AAT16518)
 CC coding for the fusion. The decorin binds to type I collagen and thus
 CC affects fibril formation. It inhibits the cell attachment-promoting
 CC activity of collagen and fibrinogen by binding to such molecules near
 CC their cell binding sites. The collagen moiety provides an integral
 CC substratum or scaffolding for the decorin. The fusion protein acts to
 CC reduce scarring of healing tissue
 XX
 XX Sequence 1107 AA;

Query Match 97.0%; Score 1149; DB 2; Length 1107;
 Best Local Similarity 93.6%; Pred. No. 7.1e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GSPGPTGLPGRGSGRPGPGADGVAGPKPAGERGSPCPA 48
 DB 297 GPPGAGEGKRGAGEPGTGLPGRGSGRPGPGADGVAGPKPAGERGSPCPA 356
 QY 49 GPKGSPGAGRPGAGLPGAGLGTGSPGSGPGDGTGTPGAGQDGRPPGPPGARGQA 108
 DB 357 GPKGSPGAGRPGAGLPGAGLGTGSPGSGPGDGTGTPGAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKGAAGSPGKAGRGVPGPGAVGAGKDGAGAGQPPGAGPAGERGEGCPA 168
 DB 417 GVMGFPKGAAGSPGKAGRGVPGPGAVGAGKDGAGAGQPPGAGPAGERGEGCPA 476
 QY 169 GSPGFQGLPGAGPGEAGKPGEGVPGDLGAPGSPGAG 208
 DB 477 GSPGFQGLPGAGPGEAGKPGEGVPGDLGAPGSPGARG 516

RESULT 12
 AAY84540
 ID AAY84540 standard; protein; 1107 AA.
 XX
 XX AAY84540;
 XX
 XX 25-JUL-2000 (first entry)
 XX
 XX Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
 XX
 XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 XX collagen; fibrinogen; fibronectin; post translational hydroxylation;
 XX decorin; chimera.
 XX
 XX Homo sapiens.
 XX Unidentified.
 XX Chimeric.
 XX
 XX Key Location/Qualifiers
 XX Misc-difference 858
 XX /note= "Gly encoded by GCT"
 XX
 XX EP992586-A2.
 XX
 XX 12-APR-2000.
 XX

PF 07-OCT-1999; 99EP-00119184.
 XX
 PR 09-OCT-1998; 98US-00169768.
 XX
 PA (USSU) US SURGICAL CORP.
 XX
 PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX
 DR WPI; 2000-259138/23.
 XX
 DR N-PSDB; AAA12500.
 XX
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 PS Claim 24; Fig 18; 260pp; English.
 CC
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention
 XX
 SQ Sequence 1107 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1107;
 Best Local Similarity 93.6%; Pred. No. 7.1e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GPPGPTGLPDPGPPGSGRGGFGADGVAGPKGPGAGERSGPGA 48
 DB 297 GPPGAGEGKRGARGEPGTGLPDPGPPGSGRGGFGADGVAGPKGPGAGERSGPGA 356
 QY 49 GPKGSPGEAGRPGAEGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 108
 DB 357 GPKGSPGEAGRPGAEGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 416
 QY 109 GVMGFPKPKGAAGEPGKAGRGVPPGAVGPKGDKGEAGAGPPGAGPAGERGQGA 168
 DB 417 GVMGFPKPKGAAGEPGKAGRGVPPGAVGPKGDKGEAGAGPPGAGPAGERGQGA 476
 QY 169 GSPFGQLPGPAGPPGAEKPGEGQGVPGDLGAPGSPGAG 208
 DB 477 GSPFGQLPGPAGPPGAEKPGEGQGVPGDLGAPGSPGAG 516

RESULT 13
 ADE87050
 ID ADE87050 standard; protein; 1161 AA.
 XX
 AC ADE87050;
 XX
 XX 29-JAN-2004 (first entry)
 XX
 XX Human pancreatic cell protein sequence SeqID510.
 XX
 XX neoplastic pancreatic cell; pancreatic cancer;
 KW cancer death; cytostatic; vaccine; gene therapy;
 XX

KW non-cancerous pancreas disease; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003060145-A2.
 XX
 PD 24-JUL-2003.
 XX
 PF 19-DEC-2002; 2002WO-US040655.
 XX
 PR 21-DEC-2001; 2001US-0342769P.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Sun Y, Liu C;
 XX
 DR WPI; 2003-587286/55.
 DR N-PSDB; ADE87387.
 XX
 PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
 PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
 PT non-cancerous disease states of the pancreas.
 XX
 PS Claim 12; SEQ ID NO 510; 635pp; English.
 CC
 CC This invention relates to novel nucleic acids and proteins present in
 CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
 CC cause of cancer death worldwide, therefore accurate methods of diagnosis
 CC and treatment are required. Compounds which modulate the proteins of the
 CC invention may have cytostatic activity and the protein and DNA sequences
 CC of the invention may be useful for the development of a vaccine or in
 CC gene therapy. The composition and methods are useful in diagnosing,
 CC staging, imaging, monitoring, preventing or treating pancreatic cancer
 CC and non-cancerous disease states of the pancreas. The present sequence is
 CC that of a human pancreatic protein of the invention.
 XX
 SQ Sequence 1161 AA;

Query Match 97.0%; Score 1149; DB 7; Length 1161;
 Best Local Similarity 93.6%; Pred. No. 7.4e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GPPGPTGLPDPGPPGSGRGGFGADGVAGPKGPGAGERSGPGA 48
 DB 155 GPPGAGEGKRGARGEPGTGLPDPGPPGSGRGGFGADGVAGPKGPGAGERSGPGA 214
 QY 49 GPKGSPGEAGRPGAEGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 108
 DB 215 GPKGSPGEAGRPGAEGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 274
 QY 109 GVMGFPKPKGAAGEPGKAGRGVPPGAVGPKGDKGEAGAGPPGAGPAGERGQGA 168
 DB 275 GVMGFPKPKGAAGEPGKAGRGVPPGAVGPKGDKGEAGAGPPGAGPAGERGQGA 334
 QY 169 GSPFGQLPGPAGPPGAEKPGEGQGVPGDLGAPGSPGAG 208
 DB 335 GSPFGQLPGPAGPPGAEKPGEGQGVPGDLGAPGSPGAG 374

RESULT 14
 AAR89469
 ID AAR89469 standard; protein; 1169 AA.
 XX
 AC AAR89469;
 XX
 XX 01-OCT-1996 (first entry)
 XX
 XX Collagen/BMP-2B fusion protein.
 XX
 KW Bone morphogenic protein 2B; BMP-2B; collagen 1A; osteogenesis;
 KW fusion protein.
 XX
 OS Synthetic.


```

XX FH Key Location/Qualifiers
XX Domain 1..1057
XX /label= "Collagen-IA"
XX /note= "collagen IA alpha-helical domain"
XX Misc-difference 887
XX /note= "unidentified amino acid"
XX Misc-difference 890
XX /note= "unidentified amino acid"
XX Peptide 1058..1059
XX /label= Linker_peptide
XX Domain 1060..1169
XX /label= BMP-2B
XX /note= "human mature BMP-2B"
XX CA2151547-A.
XX PN
XX PD 11-DEC-1995.
XX 12-JUN-1995; 95CA-02151547.
XX 10-JUN-1994; 94US-00259263.
XX (USSU ) US SURGICAL CORP.
XX Gruskin EA, Espino P;
XX WPI; 1996-140144/15.
XX N-PSDB; AAT16515.
XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
XX - and cellular regulatory factor domain, partic. useful as osteogenic
XX agents, also related vectors, transformed cells and chimaeric proteins.
XX Disclosure; Fig 5; 59pp; English.
XX A fusion protein (AAR89469) comprises the alpha-helical region of human
XX collagen I(a) linked to the human mature bone morphogenic protein 2B
XX (BMP2B). It can be expressed in Escherichia coli transformants carrying a
XX vector incorporating a chimeric gene (AAT16515) coding for the fusion.
XX The BMP moiety induces osteogenesis, while the collagen moiety provides
XX an integral substratum or scaffolding for the BMP and cells involved in
XX reconstruction and growth. The fusion protein provides sustained release
XX and delivery of BMP to a target tissue
XX Sequence 1169 AA;
Query Match 97.0%; Score 1149; DB 2; Length 1169;
Best Local Similarity 93.6%; Pred. No. 7.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTLPGPGERGSGRPFPGADGVAGPKGPAGERGSGPGA 48
DB 297 GPPGAGEGRGARGEPGTLPGPGERGSGRPFPGADGVAGPKGPAGERGSGPGA 356
QY 49 GPKSGPGEAGPGEAGLPGKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARQQA 108
DB 357 GPKSGPGEAGPGEAGLPGKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARQQA 416
QY 109 GVMGPPGKGAAGEPKKAGRGVPPGCAVAGPAGKDGAGAGQCGPPGPPGAGERGQCPA 168
DB 417 GVMGPPGKGAAGEPKKAGRGVPPGCAVAGPAGKDGAGAGQCGPPGPPGAGERGQCPA 476
QY 169 GSPGQGLPGPAGPGEAGKPGCEQGVPGDLGAPGSPGAG 208
DB 477 GSPGQGLPGPAGPGEAGKPGCEQGVPGDLGAPGSPGAG 516
RESULT 15
ID AAY84537
XX AAY84537 standard; protein; 1169 AA.
AC AAY84537;

```

```

XX DT 25-JUL-2000 (first entry)
XX DE Amino acid sequence of a chimeric collagen 1 (alpha1)/BMP-2B protein.
XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX collagen; fibrinogen; fibronectin; post translational hydroxylation;
XX ss. bone morphogenic protein; BMP-2B; chimera.
XX OS Homo sapiens.
XX OS Unidentified.
XX OS Chimeric.
XX FH Key Location/Qualifiers
XX Misc-difference 677 /note= "Ala encoded by G"
XX Misc-difference 887 /note= "unspecified amino acid encoded by CT"
XX Misc-difference 890 /note= "unspecified amino acid encoded by CT"
XX EP992586-A2.
XX 12-APR-2000.
XX 07-OCT-1999; 99EP-00119184.
XX 09-OCT-1998; 98US-00169768.
XX (USSU ) US SURGICAL CORP.
XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX WPI; 2000-259138/23.
XX N-PSDB; AAT12497.
XX Production of extracellular matrix proteins containing 4-trans-
XX hydroxyproline results in native self aggregating proteins, useful on
XX medical implants.
XX Claim 22; Fig 13; 260pp; English.
XX The specification describes a method for producing an extracellular
XX matrix protein or its fragment. The extracellular matrix protein is
XX capable of self aggregating in a cell which does not ordinarily
XX hydroxylated prolines. The method comprises optimising a nucleic acid
XX sequence for expression in the cell by substitution of codons preferred
XX by that cell for naturally occurring codons not preferred by the cell;
XX incorporating the nucleic acid sequence into the cell; and contacting the
XX cell with a hypertonic growth medium containing at least one amino acid,
XX selected from the group consisting of trans-4-hydroxyproline and 3-
XX hydroxyproline to allow at least one of the amino acids to be assimilated
XX into the cell and incorporated into the extracellular matrix protein. The
XX method may be used to make host cells assimilate and incorporate trans-4-
XX hydroxyproline into proteins. This is especially useful in the
XX recombinant production of proteins such as collagen, fibrinogen and
XX fibronectin whose ability to self aggregate and produce functional
XX proteins depends on the post translational hydroxylation of proline. The
XX method is also useful in studying the structure and function of
XX polypeptides which do not normally contain trans-4-hydroxyproline. The
XX present sequence represents a chimeric collagen 1 (alpha1)/bone
XX morphogenic protein-2B (bmp-2b) protein, which may be produced using the
XX method of the invention
XX Sequence 1169 AA;
Query Match 97.0%; Score 1149; DB 3; Length 1169;
Best Local Similarity 93.6%; Pred. No. 7.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTLPGPGERGSGRPFPGADGVAGPKGPAGERGSGPGA 48

```


Db	297	GPPGAGEGKRGEGFTGLPGPPGERGGFGSRGFPQADGVAGKGPAGERGSPGPA	356
Qy	49	GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGPAGQDGRPGPPGPARGQA	108
Db	357	GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGPAGQDGRPGPPGPARGQA	416
Qy	109	GVMGFPGPKGAAEPGKAGERGVPPGAVGPAGKDGEGAGAGQPPGPAGPAGERGEGGPA	168
Db	417	GVMGFPGPKGAAEPGKAGERGVPPGAVGPAGKDGEGAGAGQPPGPAGPAGERGEGGPA	476
Qy	169	GSPGFQGLPGPAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG	208
Db	477	GSPGFQGLPGPAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG	516

Search completed: June 17, 2005, 15:08:15
Job time : 37.1497 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:46:04 ; Search time 8.87129 Seconds
(without alignments)
1758.668 Million cell updates/sec

Title: US-10-658-989A-1

Perfect score: 1184

Sequence: 1 GPGCPGPTGLPGPPGRRG.....GEQGVGDLGNPGSPGAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1149	97.0	1341	3	US-08-963-825-18
2	1149	97.0	1341	3	US-09-500-811-18
3	1149	97.0	1341	3	US-09-570-573-18
4	1149	97.0	1341	3	US-09-548-608-18
5	1149	97.0	1461	4	US-09-585-887-9
6	1149	97.0	1461	4	US-09-289-578-9
7	1149	97.0	1464	4	US-09-331-347C-21
8	1120	94.6	1057	3	US-08-931-820-1
9	1116	94.3	595	3	US-09-213-849-48
10	1116	94.3	595	3	US-09-213-849-50
11	1116	94.3	822	3	US-09-213-849-49
12	908	76.7	492	4	US-08-468-996-12
13	866	73.1	1017	4	US-08-468-996-10
14	862	72.8	1060	3	US-08-931-820-3
15	862	72.8	1418	3	US-08-963-825-20
16	862	72.8	1418	3	US-09-010-999-1
17	862	72.8	1418	3	US-09-500-811-20
18	862	72.8	1418	3	US-09-570-573-20
19	862	72.8	1418	3	US-09-548-608-20
20	839	70.9	1442	2	US-08-316-650-12
21	839	70.9	1442	5	PCT-US95-02251-12
22	776	65.5	1057	3	US-08-931-820-4
23	765	64.6	492	4	US-08-468-996-11
24	755	63.8	1078	3	US-08-963-825-21
25	755	63.8	1078	3	US-09-500-811-21
26	755	63.8	1078	3	US-09-570-573-21
27	755	63.8	1078	3	US-09-548-608-21

Sequence 10, Appl
Sequence 10, Appl
Sequence 5882, Ap
Sequence 2, Appl
Sequence 19, Appl
Sequence 19, Appl
Sequence 19, Appl
Sequence 19, Appl
Sequence 19, Appl
Sequence 32, Appl
Sequence 65, Appl
Sequence 65, Appl
Sequence 65, Appl
Sequence 3, Appl
Sequence 52, Appl
Sequence 51, Appl
Sequence 4, Appl
Sequence 53, Appl
Sequence 62, Appl

28 745 62.9 1366 4 US-09-585-887-10
29 745 62.9 1366 4 US-09-289-578-10
30 745 62.9 1366 4 US-09-949-016-5882
31 744 62.8 1024 3 US-08-931-820-2
32 744 62.8 1366 3 US-08-963-825-19
33 744 62.8 1366 3 US-09-500-811-19
34 744 62.8 1366 3 US-09-570-573-19
35 744 62.8 1366 3 US-09-548-608-19
36 657 55.5 330 1 US-08-642-255-32
37 657 55.5 408 1 US-07-609-716-65
38 657 55.5 408 3 US-08-475-411A-65
39 657 55.5 408 3 US-08-478-029A-65
40 655.5 55.4 504 3 US-09-219-849-3
41 655.5 55.4 561 1 US-08-642-255-52
42 650.5 54.9 234 1 US-08-642-255-51
43 648 54.7 720 3 US-09-219-849-4
44 648 54.7 777 1 US-08-642-255-53
45 646.5 54.6 1064 1 US-08-642-255-62

ALIGNMENTS

RESULT 1

US-08-963-825-18

; Sequence 18, Application US/08963825

; Patent No. 6110689

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/963,825

; FILING DATE:

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/187,319

; FILING DATE: 21-JAN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Gogoris, Adda C

; REGISTRATION NUMBER: 29,714

; REFERENCE/DOCKET NUMBER: 4305/08701

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1341 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: COLLAGEN ALPHA 1 (I)

US-08-963-825-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;
 Best Local Similarity 93.6%; Pred. No. 2.9e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGSGRFPAGDVGAGPKGAGRGSPGPA 48
 |||||
 Db 334 GPPGAGEGKRGARGGEGTGLPGPPGGRGSGRFPAGDVGAGPKGAGRGSPGPA 393

QY 49 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 108
 |||||
 Db 394 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 453

QY 109 GVMGFPKGAAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 168
 |||||
 Db 454 GVMGFPKGAAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 513

QY 169 GSPGFGQLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 208
 |||||
 Db 514 GSPGFGQLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 553

RESULT 2
 US-09-500-811-18
 ; Sequence 18, Application US/09500811
 ; Patent No. 6323314
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/500,811
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1341 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (I)
 ; US-09-500-811-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;
 Best Local Similarity 93.6%; Pred. No. 2.9e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGSGRFPAGDVGAGPKGAGRGSPGPA 48
 |||||
 Db 334 GPPGAGEGKRGARGGEGTGLPGPPGGRGSGRFPAGDVGAGPKGAGRGSPGPA 393

QY 49 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 108
 |||||
 Db 394 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 453

QY 109 GVMGFPKGAAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 168
 |||||
 Db 454 GVMGFPKGAAGRPAGLPGAKGLTGTSPGSPGDKTGPAGDGRPGPPGARGOA 513

QY 169 GSPGFGQLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 208
 |||||
 Db 514 GSPGFGQLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 553

RESULT 3
 US-09-570-573-18
 ; Sequence 18, Application US/09570573
 ; Patent No. 6342361
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/570,573
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1341 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (I)
 ; US-09-570-573-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;

Best Local Similarity 93.6%; Pred. No. 2.9e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
Qy 1 GPP-----GEPGPTGLPGPGERGSGRGGPAGDGVAGPKGAGERGSPGPA 48
Db 334 GPPGAGEGKRGAGEGPGTGLPGPGERGSGRGGPAGDGVAGPKGAGERGSPGPA 393
Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTGTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 108
Db 394 GPKSGPGEAGRPGEAGLPGAKGLTGTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 453
Qy 109 GVMGFPKGAAGRPKAGRGVPGPCAVGAPGAGKDGAGAGQGGPAGPAGERGEOGPA 168
Db 454 GVMGFPKGAAGRPKAGRGVPGPCAVGAPGAGKDGAGAGQGGPAGPAGERGEOGPA 513
Qy 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
Db 514 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGARG 553

RESULT 4
US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-548-608-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2.9e-72;

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
Qy 1 GPP-----GEPGPTGLPGPGERGSGRGGPAGDGVAGPKGAGERGSPGPA 48
Db 334 GPPGAGEGKRGAGEGPGTGLPGPGERGSGRGGPAGDGVAGPKGAGERGSPGPA 393
Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTGTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 108
Db 394 GPKSGPGEAGRPGEAGLPGAKGLTGTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 453
Qy 109 GVMGFPKGAAGRPKAGRGVPGPCAVGAPGAGKDGAGAGQGGPAGPAGERGEOGPA 168
Db 454 GVMGFPKGAAGRPKAGRGVPGPCAVGAPGAGKDGAGAGQGGPAGPAGERGEOGPA 513
Qy 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
Db 514 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGARG 553

RESULT 5
US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 97.0%; Score 1149; DB 4; Length 1461;
Best Local Similarity 93.6%; Pred. No. 3.1e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
Qy 1 GPP-----GEPGPTGLPGPGERGSGRGGPAGDGVAGPKGAGERGSPGPA 48
Db 455 GPPGAGEGKRGAGEGPGTGLPGPGERGSGRGGPAGDGVAGPKGAGERGSPGPA 514
Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTGTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 108
Db 515 GPKSGPGEAGRPGEAGLPGAKGLTGTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 574
Qy 109 GVMGFPKGAAGRPKAGRGVPGPCAVGAPGAGKDGAGAGQGGPAGPAGERGEOGPA 168
Db 575 GVMGFPKGAAGRPKAGRGVPGPCAVGAPGAGKDGAGAGQGGPAGPAGERGEOGPA 634
Qy 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
Db 635 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGARG 674

RESULT 6
US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:

APPLICANT: Olsen, David R
APPLICANT: Chang, Robert
APPLICANT: McMullin, Hugh
APPLICANT: Hitzeman, Ronald A.
APPLICANT: Chisholm, George
TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
CELLS
FILE REFERENCE: 225002030400
CURRENT APPLICATION NUMBER: US/09/289, 578
PRIOR FILING DATE: 1999-04-10
PRIOR APPLICATION NUMBER: 60/084,828
PRIOR FILING DATE: 1998-05-08
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 1461
TYPE: PRT
ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 97.0%; Score 1149; DB 4; Length 1461;
Best Local Similarity 93.6%; Pred. No. 3.1e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPDPGRRGGPSRGGFGADGVAGPKGPAGERGSPGPA 48
DB 455 GPPGAGEGKRGARGEPGPTGLPDPGRRGGPSRGGFGADGVAGPKGPAGERGSPGPA 514
QY 49 GPKGSPGAGRPCEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRPPGPPGARGQA 108
DB 515 GPKGSPGAGRPCEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRPPGPPGARGQA 574
QY 109 GVMGFPKGAAGEPKGAGRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAG 168
DB 575 GVMGFPKGAAGEPKGAGRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAG 634
QY 169 GSPFGQGLPGPAGPGEAGKRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAG 208
DB 635 GSPFGQGLPGPAGPGEAGKRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAG 674

RESULT 7
US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: Obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-331-347C-21

Query Match 97.0%; Score 1149; DB 4; Length 1464;
Best Local Similarity 93.6%; Pred. No. 3.1e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPDPGRRGGPSRGGFGADGVAGPKGPAGERGSPGPA 48
DB 458 GPPGAGEGKRGARGEPGPTGLPDPGRRGGPSRGGFGADGVAGPKGPAGERGSPGPA 517
QY 49 GPKGSPGAGRPCEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRPPGPPGARGQA 108
DB 518 GPKGSPGAGRPCEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRPPGPPGARGQA 577

QY 109 GVMGFPKGAAGEPKGAGRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAG 168
DB 578 GVMGFPKGAAGEPKGAGRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAG 637
QY 169 GSPFGQGLPGPAGPGEAGKRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAG 208
DB 638 GSPFGQGLPGPAGPGEAGKRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAG 677

RESULT 8
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 94.6%; Score 1120; DB 3; Length 1057;
Best Local Similarity 96.2%; Pred. No. 2.4e-70;
Matches 200; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPDPGRRGGPSRGGFGADGVAGPKGPAGERGSPGAGKSGFGEAGRP 60
DB 309 GPNGEAGSAGPPGPPGLRGGFGSRGFGADGVAGPKGPAGERGSPGAGKSGFGEAGRP 368
QY 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRPPGPPGARGAGVGMGPPGPKGAA 120
DB 369 GEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRPPGPPGARGAGVGMGPPGPKGAA 428
QY 121 GEPKAGRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 429 GEPKAGRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
QY 181 GPGEAGKRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 208
DB 489 GPGEAGKRGVPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 516

RESULT 9
US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUNSTRA, JAN B.

; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHELLE D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 48
 ; LENGTH: 595
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-48

Query Match 94.3%; Score 1116; DB 3; Length 595;
 Best Local Similarity 90.5%; Pred. No. 2.7e-70;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGERGSGRFPFGADGVAGPKPAGERGSPGA 48
 Db 279 GPPGAGEGKRGARGEPFSGPLPGPPGERGGPGSRGFPFGADGVAGPKPAGERGAPGA 338
 Qy 49 GPKSGPGEAGRPGEAGLPGAKLGTSGSPGPDGKTGPPGAGODGRRPGRPARGQA 108
 Db 339 GPKSGPGEAGRPGEAGLPGAKLGTSGSPGPDGKTGPPGAGODGRRPGRPARGQA 398
 Qy 109 GVMGFPKGAAGPCKAGRGVPPGCAVGPAGKDGGAQAQGGPPGAPAGERGEOGPA 168
 Db 399 GVMGFPKGTAGEPKAGRGVPPGCAVGPAGKDGGAQAQGGPPGAPAGERGEOGPA 458
 Qy 169 GSPGFQGLPGPAGPGEAGKPGEGVQVFDLGAQPGSPGAG 208
 Db 459 GSPGFQGLPGPAGPGEAGKPGEGVQVFDLGAQPGSPGAG 498

RESULT 10
 US-09-219-849-50
 ; Sequence 50, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOWMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHELLE D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 49
 ; LENGTH: 822
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-49

Query Match 94.3%; Score 1116; DB 3; Length 595;
 Best Local Similarity 90.5%; Pred. No. 2.7e-70;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGERGSGRFPFGADGVAGPKPAGERGSPGA 48
 Db 279 GPPGAGEGKRGARGEPFSGPLPGPPGERGGPGSRGFPFGADGVAGPKPAGERGAPGA 338
 Qy 49 GPKSGPGEAGRPGEAGLPGAKLGTSGSPGPDGKTGPPGAGODGRRPGRPARGQA 108
 Db 339 GPKSGPGEAGRPGEAGLPGAKLGTSGSPGPDGKTGPPGAGODGRRPGRPARGQA 398
 Qy 109 GVMGFPKGAAGPCKAGRGVPPGCAVGPAGKDGGAQAQGGPPGAPAGERGEOGPA 168
 Db 399 GVMGFPKGTAGEPKAGRGVPPGCAVGPAGKDGGAQAQGGPPGAPAGERGEOGPA 458
 Qy 169 GSPGFQGLPGPAGPGEAGKPGEGVQVFDLGAQPGSPGAG 208
 Db 459 GSPGFQGLPGPAGPGEAGKPGEGVQVFDLGAQPGSPGAG 498

RESULT 11
 US-09-219-849-49
 ; Sequence 49, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOWMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHELLE D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 49
 ; LENGTH: 822
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-49

Query Match 94.3%; Score 1116; DB 3; Length 822;
 Best Local Similarity 90.5%; Pred. No. 3.6e-70;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGERGSGRFPFGADGVAGPKPAGERGSPGA 48
 Db 279 GPPGAGEGKRGARGEPFSGPLPGPPGERGGPGSRGFPFGADGVAGPKPAGERGAPGA 338
 Qy 49 GPKSGPGEAGRPGEAGLPGAKLGTSGSPGPDGKTGPPGAGODGRRPGRPARGQA 108
 Db 339 GPKSGPGEAGRPGEAGLPGAKLGTSGSPGPDGKTGPPGAGODGRRPGRPARGQA 398
 Qy 109 GVMGFPKGAAGPCKAGRGVPPGCAVGPAGKDGGAQAQGGPPGAPAGERGEOGPA 168
 Db 399 GVMGFPKGTAGEPKAGRGVPPGCAVGPAGKDGGAQAQGGPPGAPAGERGEOGPA 458
 Qy 169 GSPGFQGLPGPAGPGEAGKPGEGVQVFDLGAQPGSPGAG 208
 Db 459 GSPGFQGLPGPAGPGEAGKPGEGVQVFDLGAQPGSPGAG 498

RESULT 12


```

; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-468-996-10

Query Match 73.1%; Score 866; DB 4; Length 1017;
Best Local Similarity 74.5%; Pred. No. 9.3e-53;
Matches 155; Conservative 11; Mismatches 42; Indels 0; Gaps 0

QY 1 GPPGEPGCTGLPGRGGRGSGRGGPGADGVAGPKGAPAGERGSPGAGPKGSPGEACGRP 60
Db 292 GARGEGGVGPIPPGGRGAPGNRGFFPQQLAGPKGAPGERGSPGLAGPKGANGDGRP 351
QY 61 GEAGLPGAAGLTCSPGSPGPGDKTGTGPPGAGQDGRPGPPGPGARGQAGVGMFFGPKGAA 120
Db 352 GEPGLPGARLTCRPGDAGQGVKGVPSGAPGEDGRPGPPGQARGQPGVGMFFGPKGAN 411
QY 121 GEPKAGERVPPGPGAVGAGKDGAGAGOGPPGAPAGERGEGGAGSGFGQGLPGPA 180
Db 412 GEPKAGEKGLPGALRGLRPGKDXGEGTGAGGPPGAPAGERGEGGAGSGFGQGLPGPP 471
QY 181 GPPGEAGKPGEGQVPGDLGAPGSPGAG 208
Db 472 GPPGEAGKPGDQGVPGGAGAPGLVGRG 499

RESULT 14
US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/931,820
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO. 3:

```



```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
US-08-931-820-3

Query Match          72.8%; Score 862; DB 3; Length 1060;
Best Local Similarity 74.0%; Pred. No. 1.8e-52;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;

QY 1 GPPGEPGTGLPGPPGRRGSGRFPAGDVGAGPKGPAAGRGSPGAPGKSPGCEAGRP 60
DB 311 GARGEPGVGPIGPPGERGAPGNRFPQDGLAGPKGAPGERGSLAGPKGANGDPGRP 370
QY 61 GEAGLPKAGLTGSPGSPGPGKTPGPPAGGODGRPPGPPGARGQAGVMGFPKPGKAA 120
DB 371 GEPGLPGARGLTGRPDAGPQGVGSPGAPGEDGRPPGQARGQPGVMGFPKPGKAN 430
QY 121 GEPKAGRGVPGPGCAVGPAGKDEAGACQAGPPGAPGERGEOGPAGSPGFOGLPGPA 180
DB 431 GEPGKAGEKGLPGALRGLPKDGETGAEGLPGPAGPAGERGEOGAPGSPGFOGLPGPP 490
QY 181 GPPGAGKPGEOGVPGLDAGPSPGAPG 208
DB 491 GPPGEGKPGDQGVGEGAGAPGLVGRPG 518

RESULT 15
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; INFORMATION FOR SEQ ID NO: 20:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match          72.8%; Score 862; DB 3; Length 1418;
Best Local Similarity 74.0%; Pred. No. 2.4e-52;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;

QY 1 GPPGEPGTGLPGPPGRRGSGRFPAGDVGAGPKGPAAGRGSPGAPGKSPGCEAGRP 60
DB 423 GARGEPGVGPIGPPGERGAPGNRFPQDGLAGPKGAPGERGSLAGPKGANGDPGRP 482
QY 61 GEAGLPKAGLTGSPGSPGPGKTPGPPAGGODGRPPGPPGARGQAGVMGFPKPGKAA 120
DB 483 GEPGLPGARGLTGRPDAGPQGVGSPGAPGEDGRPPGQARGQPGVMGFPKPGKAN 542
QY 121 GEPKAGRGVPGPGCAVGPAGKDEAGACQAGPPGAPGERGEOGPAGSPGFOGLPGPA 180
DB 543 GEPKAGEKGLPGALRGLPKDGETGAEGLPGPAGPAGERGEOGAPGSPGFOGLPGPP 602
QY 181 GPPGAGKPGEOGVPGLDAGPSPGAPG 208
DB 603 GPPGEGKPGDQGVGEGAGAPGLVGRPG 630

Search completed: June 17, 2005, 15:16:48
Job time : 9.87129 secs
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 15:02:42 ; Search time 22.8937 Seconds
(without alignments)
3505.413 Million cell updates/sec

Title: US-10-658-989A-1
Perfect score: 1184
Sequence: 1 GPPGPGTGLPGRGGRGG.....GEGVPGDLGNRPSGPGAG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1714042 seqs, 38379560 residues

Total number of hits satisfying chosen parameters: 1714042

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1184	100.0	209	17	US-10-658-989A-1
2	1178	99.5	617	17	US-10-658-989A-2
3	1178	99.5	821	17	US-10-658-989A-3
4	1149	97.0	501	14	US-10-232-175-17
5	1149	97.0	1014	17	US-10-901-816A-5
6	1149	97.0	1014	17	US-10-901-816A-6
7	1149	97.0	1014	17	US-10-901-816A-9
8	1149	97.0	1014	17	US-10-901-816A-10
9	1149	97.0	1014	17	US-10-901-816A-11
10	1149	97.0	1057	15	US-10-104-889-16
11	1149	97.0	1057	15	US-10-104-889-20

12	1149	97.0	1107	15	US-10-104-889-11	Sequence 11, Appl
13	1149	97.0	1169	15	US-10-104-889-6	Sequence 8, Appl
14	1149	97.0	1171	15	US-10-104-889-8	Sequence 6, Appl
15	1149	97.0	1341	14	US-10-058-124-18	Sequence 18, Appl
16	1149	97.0	1388	15	US-10-104-889-10	Sequence 10, Appl
17	1149	97.0	1461	16	US-10-468-091-25	Sequence 25, Appl
18	1149	97.0	1464	10	US-09-918-715-261	Sequence 261, App
19	1149	97.0	1464	14	US-10-060-036-159	Sequence 159, App
20	1149	97.0	1464	14	US-10-171-311-36	Sequence 36, Appl
21	1149	97.0	1464	14	US-10-216-705-21	Sequence 21, Appl
22	1149	97.0	1464	14	US-10-149-352-2	Sequence 2, Appl
23	1149	97.0	1464	14	US-10-177-293-65	Sequence 65, Appl
24	1149	97.0	1464	14	US-10-301-822-28	Sequence 28, Appl
25	1149	97.0	1464	15	US-10-291-265-243	Sequence 243, App
26	1149	97.0	1464	16	US-10-357-851-1	Sequence 1, Appl
27	1149	97.0	1464	16	US-10-358-024-1	Sequence 1, Appl
28	1149	97.0	1464	16	US-10-734-564-79	Sequence 79, Appl
29	1149	97.0	1464	16	US-10-788-792-150	Sequence 150, App
30	1149	97.0	1464	16	US-10-474-794-261	Sequence 261, App
31	1149	97.0	1464	16	US-10-723-860-2289	Sequence 2289, Ap
32	1149	97.0	1464	17	US-10-852-335A-157	Sequence 157, App
33	1144	96.6	1449	15	US-10-402-089-8	Sequence 8, Appl
34	1144	96.6	1449	15	US-10-402-072A-8	Sequence 8, Appl
35	1141	96.4	1463	15	US-10-402-089-2	Sequence 2, Appl
36	1141	96.4	1463	15	US-10-402-072A-2	Sequence 2, Appl
37	1136	95.9	1014	17	US-10-901-816A-7	Sequence 7, Appl
38	1136	95.9	1014	17	US-10-901-816A-8	Sequence 8, Appl
39	1136	95.9	1014	17	US-10-901-816A-13	Sequence 13, Appl
40	1131	95.5	544	17	US-10-658-989A-4	Sequence 4, Appl
41	1116	94.3	595	15	US-10-342-331-48	Sequence 48, Appl
42	1116	94.3	595	15	US-10-342-331-50	Sequence 50, Appl
43	1116	94.3	822	15	US-10-342-331-49	Sequence 49, Appl
44	1116	94.3	1453	16	US-10-468-091-26	Sequence 26, Appl
45	1104	93.2	1014	17	US-10-901-816A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-10-658-989A-1
; Sequence 1, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-1
US-10-658-989A-1

Query Match	100.0%	Score 1184;	DB 17;	Length 209;
Best Local Similarity	100.0%;	Pred. No. 7.9e-59;		
Matches: 209;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	GPPGPGTGLPGRGGRGGPFCA	DVAGVAGKPGAGRGSPGAPKGS	PGEAGRP 60
Db	1	GPPGPGTGLPGRGGRGGPFCA	DVAGVAGKPGAGRGSPGAPKGS	PGEAGRP 60
Qy	61	GEAGLPGAKGLTGSFGSPGPKTK	TPGPGACQDGRPGPPGPGQGVG	FPKCAA 120


```

Db      61  GEAGLPKAGLGTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFGPKGAA 120
QY      121  GEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 180
Db      121  GEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 180
QY      181  GPPGKAGTGEQGVPPGDLGAPGSPGAG 209
Db      181  GPPGKAGTGEQGVPPGDLGAPGSPGAG 209

RESULT 2
US-10-658-989A-2
; Sequence 2, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-3
US-10-658-989A-2

Query Match      99.5%; Score 1178; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 3.9e-58;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GPPGEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAGPKGSPGAGRP 60
Db      1  GPPGEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAGPKGSPGAGRP 60
QY      61  GEAGLPKAGLGTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFGPKGAA 120
Db      61  GEAGLPKAGLGTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFGPKGAA 120
QY      121  GEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 180
Db      121  GEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 180
QY      181  GPPGKAGTGEQGVPPGDLGAPGSPGAG 208
Db      181  GPPGKAGTGEQGVPPGDLGAPGSPGAG 208

RESULT 3
US-10-658-989A-3
; Sequence 3, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3

```

```

; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-4
US-10-658-989A-3

Query Match      99.5%; Score 1178; DB 17; Length 821;
Best Local Similarity 100.0%; Pred. No. 4.9e-58;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GPPGEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAGPKGSPGAGRP 60
Db      1  GPPGEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAGPKGSPGAGRP 60
QY      61  GEAGLPKAGLGTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFGPKGAA 120
Db      61  GEAGLPKAGLGTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFGPKGAA 120
QY      121  GEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 180
Db      121  GEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 180
QY      181  GPPGKAGTGEQGVPPGDLGAPGSPGAG 208
Db      181  GPPGKAGTGEQGVPPGDLGAPGSPGAG 208

RESULT 4
US-10-232-175-17
; Sequence 17, Application US/10232175
; Publication No. US20030064074A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Olsen, David R.
; APPLICANT: Polarek, James W.
; TITLE OF INVENTION: RECOMBINANT GELATINS IN VACCINES
; FILE REFERENCE: F00224
; CURRENT APPLICATION NUMBER: US/10/232,175
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/710,249
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 501
; TYPE: PRT
; ORGANISM: human
US-10-232-175-17

Query Match      97.0%; Score 1149; DB 14; Length 501;
Best Local Similarity 93.6%; Pred. No. 1.4e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY      1  GPP-----GEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAG 48
Db      280  GPPGAGRGKRGARGEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAG 339
QY      49  GPKGSPGAGRGKRGARGEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAG 108
Db      340  GPKGSPGAGRGKRGARGEPGPTGLPGRGPGSGRFGADGVAGPKGAPGAGRGSPGAG 399
QY      109  GVMGFGPKGAGGEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 168
Db      400  GVMGFGPKGAGGEPGKAGERGVPPGPAVAGKDGAGAGQGGPAGRGAGRGQAGPAGSPGFGQLPGPA 459
QY      169  GSPGFGQLPGPAGRGKAGERGVPPGDLGAPGSPGAG 208
Db      460  GSPGFGQLPGPAGRGKAGERGVPPGDLGAPGSPGAG 499

```



```
RESULT 5
US-10-901-816A-5
; Sequence 5, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-5

Query Match          97.0%; Score 1149; DB 17; Length 1014;
Best Local Similarity 93.6%; Pred. No. 2.3e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 48
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 339

Qy 49 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 108
Db 340 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 399

Qy 109 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 168
Db 400 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 459

Qy 169 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 208
Db 460 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 499

Qy 1 GPP-----GEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 48
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 339

Qy 49 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 108
Db 340 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 399

Qy 109 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 168
Db 400 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 459

Qy 169 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 208
Db 460 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 499

RESULT 6
US-10-901-816A-6
; Sequence 6, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-6

Query Match          97.0%; Score 1149; DB 17; Length 1014;
Best Local Similarity 93.6%; Pred. No. 2.3e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 48
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 339

Qy 49 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 108
Db 340 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 399

Qy 109 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 168
Db 400 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 459

Qy 169 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 208
Db 460 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 499

RESULT 7
US-10-901-816A-9
; Sequence 9, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-9

Query Match          97.0%; Score 1149; DB 17; Length 1014;
Best Local Similarity 93.6%; Pred. No. 2.3e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 48
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 339

Qy 49 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 108
Db 340 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 399

Qy 109 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 168
Db 400 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 459

Qy 169 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 208
Db 460 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 499

RESULT 8
US-10-901-816A-10
; Sequence 10, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
```

```
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 339

Qy 49 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 108
Db 340 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 399

Qy 109 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 168
Db 400 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 459

Qy 169 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 208
Db 460 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 499

RESULT 7
US-10-901-816A-9
; Sequence 9, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-9

Query Match          97.0%; Score 1149; DB 17; Length 1014;
Best Local Similarity 93.6%; Pred. No. 2.3e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 48
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGCGSRGPPGADGVAGPKPAGERGSPGA 339

Qy 49 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 108
Db 340 GPKGSPGAGRPGEAGLPGAKGLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQA 399

Qy 109 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 168
Db 400 GVMGFPKGAAGEPGKAGRGVPPGCAVGPAGKDGAGAQGGPPGPPGAGERGEOGPA 459

Qy 169 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 208
Db 460 GSPGFQGLPGPAGPPGAGKPGEGVPGDGLGAPGSPGAG 499

RESULT 8
US-10-901-816A-10
; Sequence 10, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
```


; PRIOR FILING DATE: 2003-08-01
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 10
 ; LENGTH: 1014
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-901-816A-10

Query Match 97.0%; Score 1149; DB 17; Length 1014;
 Best Local Similarity 93.6%; Pred. No. 2.3e-56;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY	1	GPP-----GEPGPTGLPGPGRGPGSRGFPAGDVGAGKPGAGRGSPGPA	48
DB	280	GPPGAGEGKRGARGEPGPTGLPGPGRGPGSRGFPAGDVGAGKPGAGRGSPGPA	339
QY	49	GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPRGAGDGRPGPPPGARGQA	108
DB	340	GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPRGAGDGRPGPPPGARGQA	399
QY	109	GVMGFPKGKGAAGPFGKAGRGVFPFGCAVGPAGKDGAGAGQPPGPPGAGRGSPGPA	168
DB	400	GVMGFPKGKGAAGPFGKAGRGVFPFGCAVGPAGKDGAGAGQPPGPPGAGRGSPGPA	459
QY	169	GSPGFQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG	208
DB	460	GSPGFQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG	499

RESULT 9
 US-10-901-816A-11
 ; Sequence 11, Application US/10901816A
 ; Publication No. US20050056703A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Robert C.
 ; APPLICANT: Olsen, David R.
 ; APPLICANT: James, Polarek W.
 ; APPLICANT: Williams, Kim E.
 ; TITLE OF INVENTION: Gelatin Capsules
 ; FILE REFERENCE: FP0404 US
 ; CURRENT APPLICATION NUMBER: US/10/901,816A
 ; CURRENT FILING DATE: 2004-07-29
 ; PRIOR APPLICATION NUMBER: US 60/492,085
 ; PRIOR FILING DATE: 2003-08-01
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 11
 ; LENGTH: 1014
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-901-816A-11

Query Match 97.0%; Score 1149; DB 17; Length 1014;
 Best Local Similarity 93.6%; Pred. No. 2.3e-56;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY	1	GPP-----GEPGPTGLPGPGRGPGSRGFPAGDVGAGKPGAGRGSPGPA	48
DB	280	GPPGAGEGKRGARGEPGPTGLPGPGRGPGSRGFPAGDVGAGKPGAGRGSPGPA	339
QY	49	GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPRGAGDGRPGPPPGARGQA	108
DB	340	GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPRGAGDGRPGPPPGARGQA	399
QY	109	GVMGFPKGKGAAGPFGKAGRGVFPFGCAVGPAGKDGAGAGQPPGPPGAGRGSPGPA	168
DB	400	GVMGFPKGKGAAGPFGKAGRGVFPFGCAVGPAGKDGAGAGQPPGPPGAGRGSPGPA	459
QY	169	GSPGFQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG	208
DB	460	GSPGFQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG	499

RESULT 10
 US-10-104-889-16
 ; Sequence 16, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GRUSKIN, ELLIOT A.
 ; BUECHTER, DOUGLAS
 ; BROKAW, JANE
 ; ZHANG, GUANGHUI
 ; PAOLELLA, DAVID
 ; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DILWORTH & BARRESE
 ; STREET: 333 EARLE OVINGTON BOULEVARD
 ; CITY: UNIONDALE
 ; STATE: NY
 ; COUNTRY: U.S.A.
 ; ZIP: 11553
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/104,889
 ; FILING DATE: 22-Mar-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/169,768
 ; FILING DATE: 09-OCT-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEEN, JEFFREY S
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 228-8484
 ; TELEFAX: (516) 228-8516
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1057 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
 US-10-104-889-16

Query Match	97.0%;	Score 1149;	DB 15;	Length 1057;
Best Local Similarity	93.6%;	Pred. No. 2.4e-56;		
Matches	206;	Conservative	0;	Mismatches 2; Indels 12; Gaps 1;
QY	1	GPP-----GEPGPTGLPGPGRGPGSRGFPAGDVGAGKPGAGRGSPGPA	48	
DB	297	GPPGAGEGKRGARGEPGPTGLPGPGRGPGSRGFPAGDVGAGKPGAGRGSPGPA	356	
QY	49	GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPRGAGDGRPGPPPGARGQA	108	
DB	357	GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPRGAGDGRPGPPPGARGQA	416	
QY	109	GVMGFPKGKGAAGPFGKAGRGVFPFGCAVGPAGKDGAGAGQPPGPPGAGRGSPGPA	168	
DB	417	GVMGFPKGKGAAGPFGKAGRGVFPFGCAVGPAGKDGAGAGQPPGPPGAGRGSPGPA	476	
QY	169	GSPGFQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG	208	
DB	477	GSPGFQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG	516	

RESULT 11
 US-10-104-889-20
 ; Sequence 20, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:


```

/ COUNTRY: U.S.A.
/ ZIP: 11553
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/104,889
/ FILING DATE: 22-Mar-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/169,768
/ FILING DATE: 09-OCT-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: STEEN, JEFFREY S
/ TELEPHONE: (516) 228-8484
/ TELEFAX: (516) 228-8516
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1169 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-104-889-6

Query Match          97.0%; Score 1149; DB 15; Length 1169;
Best Local Similarity 93.6%; Pred. No. 2.6e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTLPGPPGGRGSGRFPFGADGVAGPKPAGERGSPGPA 48
|||
DB 297 GPPGAGEGKRGARGEPGTLPGPPGGRGSGRFPFGADGVAGPKPAGERGSPGPA 356

QY 49 GPKSGPEAGRPEAGLPGAKGLTSGPSGPDGKTGPPGAGODGRGPPGPGARQQA 108
|||
DB 357 GPKSGPEAGRPEAGLPGAKGLTSGPSGPDGKTGPPGAGODGRGPPGPGARQQA 416

QY 109 GWMGFPKGAAGEPKAGRGVPPGCAVGPAGKDGAGAGCPGPPGAPAGERGEOGPA 168
|||
DB 417 GWMGFPKGAAGEPKAGRGVPPGCAVGPAGKDGAGAGCPGPPGAPAGERGEOGPA 476

QY 169 GSPFGQLPGPAGPGEAGKPGEOGVPGDLGAPGSPGAG 208
|||
DB 477 GSPFGQLPGPAGPGEAGKPGEOGVPGDLGAPGSPGAG 516

RESULT 14
US-10-104-889-8
/ Sequence 8, Application US/10104889
/ Publication No. US20040086961A1
/ GENERAL INFORMATION:
/ APPLICANT: GRUSKIN, ELLIOT A.
/ BUECHTER, DOUGLAS
/ BROKAW, JANE
/ ZHANG, GUANGHUI
/ PAOLELLA, DAVID
/ TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
/ NUMBER OF SEQUENCES: 50
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: DILWORTH & BARRESE
/ STREET: 333 EARLE OVLINGTON BOULEVARD
/ CITY: UNIONDALE
/ STATE: NY
/ COUNTRY: U.S.A.
/ ZIP: 11553
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/104,889
/ FILING DATE: 22-Mar-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/169,768
/ FILING DATE: 09-OCT-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: STEEN, JEFFREY S
/ TELEPHONE: (516) 228-8484
/ TELEFAX: (516) 228-8516
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1171 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-889-8

Query Match          97.0%; Score 1149; DB 15; Length 1171;
Best Local Similarity 93.6%; Pred. No. 2.6e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GPPGTLPGPPGGRGSGRFPFGADGVAGPKPAGERGSPGPA 48
|||
DB 297 GPPGAGEGKRGARGEPGTLPGPPGGRGSGRFPFGADGVAGPKPAGERGSPGPA 356

QY 49 GPKSGPEAGRPEAGLPGAKGLTSGPSGPDGKTGPPGAGODGRGPPGPGARQQA 108
|||
DB 357 GPKSGPEAGRPEAGLPGAKGLTSGPSGPDGKTGPPGAGODGRGPPGPGARQQA 416

QY 109 GWMGFPKGAAGEPKAGRGVPPGCAVGPAGKDGAGAGCPGPPGAPAGERGEOGPA 168
|||
DB 417 GWMGFPKGAAGEPKAGRGVPPGCAVGPAGKDGAGAGCPGPPGAPAGERGEOGPA 476

QY 169 GSPFGQLPGPAGPGEAGKPGEOGVPGDLGAPGSPGAG 208
|||
DB 477 GSPFGQLPGPAGPGEAGKPGEOGVPGDLGAPGSPGAG 516

RESULT 15
US-10-058-124-18
/ Sequence 18, Application US/10058124
/ Publication No. US20030119058A1
/ GENERAL INFORMATION:
/ APPLICANT: Qvist, Per
/ Bonde, Martin
/ TITLE OF INVENTION: A Method for Assaying Collagen Fragments
/ in Body Fluids, A Test Kit and Means for Carrying Out the
/ Method and Use of the Method to Diagnose the Presence of
/ Disorders Associated with the Metabolism of
/
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Darby & Darby PC
/ STREET: 805 Third Avenue
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10022
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/058,124
/ FILING DATE: 29-Jan-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:

```


APPLICATION NUMBER: 09/570,573
 FILING DATE: 2002-MAY-12
 APPLICATION NUMBER: 08/187,319
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236687
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1341 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN ALPHA 1 (I)
 SEQUENCE DESCRIPTION: SEQ ID NO: 18:
 US-10-058-124-18

Query Match 97.0%; Score 1149; DB 14; Length 1341;
 Best Local Similarity 93.6%; Pred. No. 2.9e-56;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Qy	1	GPP-----GEPGTLGPPGGRGGSGRGGPGADGVAGPKGPGAGRGSGGPA	48
Db	334	GPPGAGEGKRGAGEPGTGLPFGPGGRGGSGRGGPGADGVAGPKGPGAGRGSGGPA	393
Qy	49	GPKGSPGEAGRPGEAGLFGAKGLTSGSGSGPGDGTGTPPGAGQDGRPPGPPGARGQA	108
Db	394	GPKGSPGEAGRPGEAGLFGAKGLTSGSGSGPGDGTGTPPGAGQDGRPPGPPGARGQA	453
Qy	109	GVMGFPGPKGAAGEPKGAGRGVPGPPGAVGPAKDGAGAGCGPPGPPGAGRGGEQGPA	168
Db	454	GVMGFPGPKGAAGEPKGAGRGVPGPPGAVGPAKDGAGAGCGPPGPPGAGRGGEQGPA	513
Qy	169	GSRGFGGLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG	208
Db	514	GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG	553

Search completed: June 17, 2005, 15:35:29
 Job time : 24.8937 secs

THIS PAGE BLANK (USPTO)

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	1149	97.0	1464	1	CGHU1S
2	1118	94.4	671	1	CGRT1S
3	1116	94.3	1453	2	S21626
4	1067	90.1	1042	1	CGGH1S
5	862	72.8	1487	1	CGHU6C
6	860	72.6	1418	2	T45467
7	850	71.8	1486	1	CGHU33
8	847	71.5	1492	2	A40333
9	839	70.9	1419	2	A41182
10	839	70.9	1487	2	B41182
11	824	69.6	779	1	CGBO1S
12	815	68.8	673	1	CGBO6C
13	794.5	67.1	886	2	I50694
14	786	66.4	1496	1	CGHU2S
15	776	65.5	1466	1	CGHU7L
16	765	64.6	1497	2	I49607
17	763	64.4	1464	2	S59856
18	761.5	64.3	1049	1	CGBO7S
19	760	64.2	1373	1	A43291
20	744	62.8	1366	1	CGHU2S
21	629.5	53.2	1027	2	S28774
22	607	51.3	1838	1	CGHU1V
23	602.5	50.9	310	2	I50696
24	600.5	50.7	1414	1	S23809
25	599	50.6	1843	2	S18803
26	590.5	49.9	632	2	S42731
27	589	49.7	964	1	CGCH2S
28	577	48.7	1806	1	CGHU1E
29	576.5	48.7	7730	2	A36226

J. Biol. Chem. 260, 2315-2320, 1985
A/Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter structure and organization of the 5' and 3' ends
A/Reference number: 155237; MUID:85130970; PMID:2857713
A/Accession: 155237
A/Status: translation not shown; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-34 <CH2>
A/Cross-references: GB:M10627; NID:g180383; PIDN:AA51992.1; PID:g553226
R/Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, J. Biol. Chem. 265, 6312-6317, 1990
A/Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-terminal propeptides of type I procollagen
A/Reference number: A35233; MUID:90202908; PMID:2318855
A/Accession: A35233
A/Molecule type: protein
A/Residues: 33-52 <WR>
A/Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved alpha 1(I) procollagen molecule
R/Wall, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F. EMBO J. 8, 1705-1710, 1989
A/Title: A base substitution in the exon of a collagen gene causes alternative splicing of the alpha 1(I) procollagen mRNA
A/Reference number: S09400; MUID:89356643; PMID:2767050
A/Accession: S09400
A/Molecule type: mRNA
A/Residues: 156-183 <WEI>
R/Click, E.M.; Bornstein, P. Biochemistry 9, 4699-4706, 1970
A/Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) procollagen
A/Reference number: A90567; MUID:71038625; PMID:5529814
A/Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
A/Accession: B90567
A/Molecule type: protein
A/Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z', 235, 'Z', 238, 'Z', 240-241, 'Z', 243-244, 'Z', 246-247, 'Z', 249-250, 'Z', 252-253, 'Z', 255-256, 'Z', 258-259, 'Z', 261-262, 'Z', 264-265, 'Z', 267-268, 'Z', 270-271, 'Z', 273-274, 'Z', 276-277, 'Z', 279-280, 'Z', 282-283, 'Z', 285-286, 'Z', 288-289, 'Z', 291-292, 'Z', 294-295, 'Z', 297-298, 'Z', 300-301, 'Z', 303-304, 'Z', 306-307, 'Z', 309-310, 'Z', 312-313, 'Z', 315-316, 'Z', 318-319, 'Z', 321-322, 'Z', 324-325, 'Z', 327-328, 'Z', 330-331, 'Z', 333-334, 'Z', 336-337, 'Z', 339-340, 'Z', 342-343, 'Z', 345-346, 'Z', 348-349, 'Z', 351-352, 'Z', 354-355, 'Z', 357-358, 'Z', 360-361, 'Z', 363-364, 'Z', 366-367, 'Z', 369-370, 'Z', 372-373, 'Z', 375-376, 'Z', 378-379, 'Z', 381-382, 'Z', 384-385, 'Z', 387-388, 'Z', 390-391, 'Z', 393-394, 'Z', 396-397, 'Z', 399-400, 'Z', 402-403, 'Z', 405-406, 'Z', 408-409, 'Z', 411-412, 'Z', 414-415, 'Z', 417-418, 'Z', 420-421, 'Z', 423-424, 'Z', 426-427, 'Z', 429-430, 'Z', 432-433, 'Z', 435-436, 'Z', 438-439, 'Z', 441-442, 'Z', 444-445, 'Z', 447-448, 'Z', 450-451, 'Z', 453-454, 'Z', 456-457, 'Z', 459-460, 'Z', 462-463, 'Z', 465-466, 'Z', 468-469, 'Z', 471-472, 'Z', 474-475, 'Z', 477-478, 'Z', 480-481, 'Z', 483-484, 'Z', 486-487, 'Z', 489-490, 'Z', 492-493, 'Z', 495-496, 'Z', 498-499, 'Z', 501-502, 'Z', 504-505, 'Z', 507-508, 'Z', 510-511, 'Z', 513-514, 'Z', 516-517, 'Z', 519-520, 'Z', 522-523, 'Z', 524-525, 'Z', 527-528, 'Z', 530-531, 'Z', 533-534, 'Z', 536-537, 'Z', 539-540, 'Z', 542-543, 'Z', 545-546, 'Z', 548-549, 'Z', 551-552, 'Z', 554-555, 'Z', 557-558, 'Z', 560-561, 'Z', 563-564, 'Z', 565-566, 'Z', 568-569, 'Z', 571-572, 'Z', 574-575, 'Z', 577-578, 'Z', 580-581, 'Z', 583-584, 'Z', 586-587, 'Z', 589-590, 'Z', 592-593, 'Z', 595-596, 'Z', 598-599, 'Z', 601-602, 'Z', 604-605, 'Z', 607-608, 'Z', 610-611, 'Z', 613-614, 'Z', 616-617, 'Z', 619-620, 'Z', 622-623, 'Z', 624-625, 'Z', 627-628, 'Z', 630-631, 'Z', 633-634, 'Z', 636-637, 'Z', 639-640, 'Z', 642-643, 'Z', 645-646, 'Z', 648-649, 'Z', 651-652, 'Z', 654-655, 'Z', 657-658, 'Z', 660-661, 'Z', 663-664, 'Z', 666-667, 'Z', 669-670, 'Z', 672-673, 'Z', 675-676, 'Z', 678-679, 'Z', 681-682, 'Z', 684-685, 'Z', 687-688, 'Z', 690-691, 'Z', 693-694, 'Z', 696-697, 'Z', 699-700, 'Z', 702-703, 'Z', 705-706, 'Z', 708-709, 'Z', 711-712, 'Z', 714-715, 'Z', 717-718, 'Z', 720-721, 'Z', 723-724, 'Z', 726-727, 'Z', 729-730, 'Z', 732-733, 'Z', 735-736, 'Z', 738-739, 'Z', 741-742, 'Z', 744-745, 'Z', 747-748, 'Z', 750-751, 'Z', 753-754, 'Z', 756-757, 'Z', 759-760, 'Z', 762-763, 'Z', 765-766, 'Z', 768-769, 'Z', 771-772, 'Z', 774-775, 'Z', 777-778, 'Z', 780-781, 'Z', 783-784, 'Z', 786-787, 'Z', 789-790, 'Z', 792-793, 'Z', 795-796, 'Z', 798-799, 'Z', 801-802, 'Z', 804-805, 'Z', 807-808, 'Z', 810-811, 'Z', 813-814, 'Z', 816-817, 'Z', 819-820, 'Z', 822-823, 'Z', 824-825, 'Z', 827-828, 'Z', 830-831, 'Z', 833-834, 'Z', 836-837, 'Z', 839-840, 'Z', 842-843, 'Z', 845-846, 'Z', 848-849, 'Z', 851-852, 'Z', 854-855, 'Z', 857-858, 'Z', 860-861, 'Z', 863-864, 'Z', 866-867, 'Z', 869-870, 'Z', 872-873, 'Z', 875-876, 'Z', 878-879, 'Z', 881-882, 'Z', 884-885, 'Z', 887-888, 'Z', 890-891, 'Z', 893-894, 'Z', 896-897, 'Z', 899-900, 'Z', 902-903, 'Z', 904-905, 'Z', 907-908, 'Z', 910-911, 'Z', 913-914, 'Z', 916-917, 'Z', 919-920, 'Z', 922-923, 'Z', 924-925, 'Z', 927-928, 'Z', 930-931, 'Z', 933-934, 'Z', 936-937, 'Z', 939-940, 'Z', 942-943, 'Z', 945-946, 'Z', 948-949, 'Z', 951-952, 'Z', 954-955, 'Z', 957-958, 'Z', 960-961, 'Z', 963-964, 'Z', 966-967, 'Z', 969-970, 'Z', 972-973, 'Z', 975-976, 'Z', 978-979, 'Z', 981-982, 'Z', 984-985, 'Z', 987-988, 'Z', 990-991, 'Z', 993-994, 'Z', 996-997, 'Z', 999-1000, 'Z', 1002-1003, 'Z', 1004-1005, 'Z', 1007-1008, 'Z', 1010-1011, 'Z', 1013-1014, 'Z', 1016-1017, 'Z', 1019-1020, 'Z', 1022-1023, 'Z', 1024-1025, 'Z', 1027-1028, 'Z', 1030-1031, 'Z', 1033-1034, 'Z', 1036-1037, 'Z', 1039-1040, 'Z', 1042-1043, 'Z', 1045-1046, 'Z', 1048-1049, 'Z', 1051-1052, 'Z', 1054-1055, 'Z', 1057-1058, 'Z', 1060-1061, 'Z', 1063-1064, 'Z', 1066-1067, 'Z', 1069-1070, 'Z', 1072-1073, 'Z', 1075-1076, 'Z', 1078-1079, 'Z', 1081-1082, 'Z', 1084-1085, 'Z', 1087-1088, 'Z', 1090-1091, 'Z', 1093-1094, 'Z', 1096-1097, 'Z', 1099-1100, 'Z', 1102-1103, 'Z', 1104-1105, 'Z', 1107-1108, 'Z', 1110-1111, 'Z', 1113-1114, 'Z', 1116-1117, 'Z', 1119-1120, 'Z', 1122-1123, 'Z', 1124-1125, 'Z', 1127-1128, 'Z', 1130-1131, 'Z', 1133-1134, 'Z', 1136-1137, 'Z', 1139-1140, 'Z', 1142-1143, 'Z', 1145-1146, 'Z', 1148-1149, 'Z', 1151-1152, 'Z', 1154-1155, 'Z', 1157-1158, 'Z', 1160-1161, 'Z', 1163-1164, 'Z', 1166-1167, 'Z', 1169-1170, 'Z', 1172-1173, 'Z', 1175-1176, 'Z', 1178-1179, 'Z', 1181-1182, 'Z', 1184-1185, 'Z', 1187-1188, 'Z', 1190-1191, 'Z', 1193-1194, 'Z', 1196-1197, 'Z', 1199-1200, 'Z', 1202-1203, 'Z', 1204-1205, 'Z', 1207-1208, 'Z', 1210-1211, 'Z', 1213-1214, 'Z', 1216-1217, 'Z', 1219-1220, 'Z', 1222-1223, 'Z', 1224-1225, 'Z', 1227-1228, 'Z', 1230-1231, 'Z', 1233-1234, 'Z', 1236-1237, 'Z', 1239-1240, 'Z', 1242-1243, 'Z', 1245-1246, 'Z', 1248-1249, 'Z', 1251-1252, 'Z', 1254-1255, 'Z', 1257-1258, 'Z', 1260-1261, 'Z', 1263-1264, 'Z', 1266-1267, 'Z', 1269-1270, 'Z', 1272-1273, 'Z', 1275-1276, 'Z', 1278-1279, 'Z', 1281-1282, 'Z', 1284-1285, 'Z', 1287-1288, 'Z', 1290-1291, 'Z', 1293-1294, 'Z', 1296-1297, 'Z', 1299-1300, 'Z', 1302-1303, 'Z', 1304-1305, 'Z', 1307-1308, 'Z', 1310-1311, 'Z', 1313-1314, 'Z', 1316-1317, 'Z', 1319-1320, 'Z', 1322-1323, 'Z', 1324-1325, 'Z', 1327-1328, 'Z', 1330-1331, 'Z', 1333-1334, 'Z', 1336-1337, 'Z', 1339-1340, 'Z', 1342-1343, 'Z', 1345-1346, 'Z', 1348-1349, 'Z', 1351-1352, 'Z', 1354-1355, 'Z', 1357-1358, 'Z', 1360-1361, 'Z', 1363-1364, 'Z', 1366-1367, 'Z', 1369-1370, 'Z', 1372-1373, 'Z', 1375-1376, 'Z', 1378-1379, 'Z', 1381-1382, 'Z', 1384-1385, 'Z', 1387-1388, 'Z', 1390-1391, 'Z', 1393-1394, 'Z', 1396-1397, 'Z', 1399-1400, 'Z', 1402-1403, 'Z', 1405-1406, 'Z', 1408-1409, 'Z', 1411-1412, 'Z', 1414-1415, 'Z', 1417-1418, 'Z', 1420-1421, 'Z', 1423-1424, 'Z', 1426-1427, 'Z', 1429-1430, 'Z', 1432-1433, 'Z', 1435-1436, 'Z', 1438-1439, 'Z', 1441-1442, 'Z', 1444-1445, 'Z', 1447-1448, 'Z', 1450-1451, 'Z', 1453-1454, 'Z', 1456-1457, 'Z', 1459-1460, 'Z', 1462-1463, 'Z', 1465-1466, 'Z', 1468-1469, 'Z', 1471-1472, 'Z', 1474-1475, 'Z', 1477-1478, 'Z', 1480-1481, 'Z', 1483-1484, 'Z', 1486-1487, 'Z', 1489-1490, 'Z', 1492-1493, 'Z', 1495-1496, 'Z', 1498-1499, 'Z', 1501-1502, 'Z', 1504-1505, 'Z', 1507-1508, 'Z', 1510-1511, 'Z', 1513-1514, 'Z', 1516-1517, 'Z', 1519-1520, 'Z', 1522-1523, 'Z', 1524-1525, 'Z', 1527-1528, 'Z', 1530-1531, 'Z', 1533-1534, 'Z', 1536-1537, 'Z', 1539-1540, 'Z', 1542-1543, 'Z', 1545-1546, 'Z', 1548-1549, 'Z', 1551-1552, 'Z', 1554-1555, 'Z', 1557-1558, 'Z', 1560-1561, 'Z', 1563-1564, 'Z', 1566-1567, 'Z', 1569-1570, 'Z', 1572-1573, 'Z', 1575-1576, 'Z', 1578-1579, 'Z', 1581-1582, 'Z', 1584-1585, 'Z', 1587-1588, 'Z', 1590-1591, 'Z', 1593-1594, 'Z', 1596-1597, 'Z', 1599-1600, 'Z', 1602-1603, 'Z', 1604-1605, 'Z', 1607-1608, 'Z', 1610-1611, 'Z', 1613-1614, 'Z', 1616-1617, 'Z', 1619-1620, 'Z', 1622-1623, 'Z', 1624-1625, 'Z', 1627-1628, 'Z', 1630-1631, 'Z', 1633-1634, 'Z', 1636-1637, 'Z', 1639-1640, 'Z', 1642-1643, 'Z', 1645-1646, 'Z', 1648-1649, 'Z', 1651-1652, 'Z', 1654-1655, 'Z', 1657-1658, 'Z', 1660-1661, 'Z', 1663-1664, 'Z', 1666-1667, 'Z', 1669-1670, 'Z', 1672-1673, 'Z', 1675-1676, 'Z', 1678-1679, 'Z', 1681-1682, 'Z', 1684-1685, 'Z', 1687-1688, 'Z', 1690-1691, 'Z', 1693-1694, 'Z', 1696-1697, 'Z', 1699-1700, 'Z', 1702-1703, 'Z', 1705-1706, 'Z', 1708-1709, 'Z', 1711-1712, 'Z', 1714-1715, 'Z', 1717-1718, 'Z', 1720-1721, 'Z', 1723-1724, 'Z', 1726-1727, 'Z', 1729-1730, 'Z', 1732-1733, 'Z', 1735-1736, 'Z', 1738-1739, 'Z', 1741-1742, 'Z', 1744-1745, 'Z', 1747-1748, 'Z', 1750-1751, 'Z', 1753-1754, 'Z', 1756-1757, 'Z', 1759-1760, 'Z', 1762-1763, 'Z', 1765-1766, 'Z', 1768-1769, 'Z', 1771-1772, 'Z', 1774-1775, 'Z', 1777-1778, 'Z', 1780-1781, 'Z', 1783-1784, 'Z', 1786-1787, 'Z', 1789-1790, 'Z', 1792-1793, 'Z', 1795-1796, 'Z', 1798-1799, 'Z', 1801-1802, 'Z', 1804-1805, 'Z', 1807-1808, 'Z', 1810-1811, 'Z', 1813-1814, 'Z', 1816-1817, 'Z', 1819-1820, 'Z', 1822-1823, 'Z', 1824-1825, 'Z', 1827-1828, 'Z', 1830-1831, 'Z', 1833-1834, 'Z', 1836-1837, 'Z', 1839-1840, 'Z', 1842-1843, 'Z', 1845-1846, 'Z', 1848-1849, 'Z', 1851-1852, 'Z', 1854-1855, 'Z', 1857-1858, 'Z', 1860-1861, 'Z', 1863-1864, 'Z', 1866-1867, 'Z', 1869-1870, 'Z', 1872-1873, 'Z', 1875-1876, 'Z', 1878-1879, 'Z', 1881-1882, 'Z', 1884-1885, 'Z', 1887-1888, 'Z', 1890-1891, 'Z', 1893-1894, 'Z', 1896-1897, 'Z', 1899-1900, 'Z', 1902-1903, 'Z', 1904-1905, 'Z', 1907-1908, 'Z', 1910-1911, 'Z', 1913-1914, 'Z', 1916-1917, 'Z', 1919-1920, 'Z', 1922-1923, 'Z', 1924-1925, 'Z', 1927-1928, 'Z', 1930-1931, 'Z', 1933-1934, 'Z', 1936-1937, 'Z', 1939-1940, 'Z', 1942-1943, 'Z', 1945-1946, 'Z', 1948-1949, 'Z', 1951-1952, 'Z', 1954-1955, 'Z', 1957-1958, 'Z', 1960-1961, 'Z', 1963-1964, 'Z', 1966-1967, 'Z', 1969-1970, 'Z', 1972-1973, 'Z', 1975-1976, 'Z', 1978-1979, 'Z', 1981-1982, 'Z', 1984-1985, 'Z', 1987-1988, 'Z', 1990-1991, 'Z', 1993-1994, 'Z', 1996-1997, 'Z', 1999-2000, 'Z', 2002-2003, 'Z', 2004-2005, 'Z', 2007-2008, 'Z', 2010-2011, 'Z', 2013-2014, 'Z', 2016-2017, 'Z', 2019-2020, 'Z', 2022-2023, 'Z', 2024-2025, 'Z', 2027-2028, 'Z', 2030-2031, 'Z', 2033-2034, 'Z', 2036-2037, 'Z', 2039-2040, 'Z', 2042-2043, 'Z', 2045-2046, 'Z', 2048-2049, 'Z', 2051-2052, 'Z', 2054-2055, 'Z', 2057-2058, 'Z', 2060-2061, 'Z', 2063-2064, 'Z', 2066-2067, 'Z', 2069-2070, 'Z', 2072-2073, 'Z', 2075-2076, 'Z', 2078-2079, 'Z', 2081-2082, 'Z', 2084-2085, 'Z', 2087-2088, 'Z', 2090-2091, 'Z', 2093-2094, 'Z', 2096-2097, 'Z', 2099-2100, 'Z', 2102-2103, 'Z', 2104-2105, 'Z', 2107-2108, 'Z', 2110-2111, 'Z', 2113-2114, 'Z', 2116-2117, 'Z', 2119-2120, 'Z', 2122-2123, 'Z', 2124-2125, 'Z', 2127-2128, 'Z', 2130-2131, 'Z', 2133-2134, 'Z', 2136-2137, 'Z', 2139-2140, 'Z', 2142-2143, 'Z', 2145-2146, 'Z', 2148-2149, 'Z', 2151-2152, 'Z', 2154-2155, 'Z', 2157-2158, 'Z', 2160-2161, 'Z', 2163-2164, 'Z', 2166-2167, 'Z', 2169-2170, 'Z', 2172-2173, 'Z', 2175-2176, 'Z', 2178-2179, 'Z', 2181-2182, 'Z', 2184-2185, 'Z', 2187-2188, 'Z', 2190-2191, 'Z', 2193-2194, 'Z', 2196-2197, 'Z', 2199-2200, 'Z', 2202-2203, 'Z', 2204-2205, 'Z', 2207-2208, 'Z', 2210-2211, 'Z', 2213-2214, 'Z', 2216-2217, 'Z', 2219-2220, 'Z', 2222-2223, 'Z', 2224-2225, 'Z', 2227-2228, 'Z', 2230-2231, 'Z', 2233-2234, 'Z', 2236-2237, 'Z', 2239-2240, 'Z', 2242-2243, 'Z', 2245-2246, 'Z', 2248-2249, 'Z', 2251-2252, 'Z', 2254-2255, 'Z', 2257-2258, 'Z', 2260-2261, 'Z', 2263-2264, 'Z', 2266-2267, 'Z', 2269-2270, 'Z', 2272-2273, 'Z', 2275-2276, 'Z', 2278-2279, 'Z', 2281-2282, 'Z', 2284-2285, 'Z', 2287-2288, 'Z', 2290-2291, 'Z', 2293-2294, 'Z', 2296-2297, 'Z', 2299-2300, 'Z', 2302-2303, 'Z', 2304-2305, 'Z', 2307-2308, 'Z', 2310-2311, 'Z', 2313-2314, 'Z', 2316-2317, 'Z', 2319-2320, 'Z', 2322-2323, 'Z', 2324-2325, 'Z', 2327-2328, 'Z', 2330-2331, 'Z', 2333-2334, 'Z', 2336-2337, 'Z', 2339-2340, 'Z', 2342-2343, 'Z', 2345-2346, 'Z', 2348-2349, 'Z', 2351-2352, 'Z', 2354-2355, 'Z', 2357-2358, 'Z', 2360-2361, 'Z', 2363-2364, 'Z', 2366-2367, 'Z', 2369-2370, 'Z', 2372-2373, 'Z', 2375-2376, 'Z', 2378-2379, 'Z', 2381-2382, 'Z', 2384-2385, 'Z', 2387-2388, 'Z', 2390-2391, 'Z', 2393-2394, 'Z', 2396-2397, 'Z', 2399-2400, 'Z', 2402-2403, 'Z', 2405-2406, 'Z', 2408-2409, 'Z', 2411-2412, 'Z', 2414-2415, 'Z', 2417-2418, 'Z', 2420-2421, 'Z', 2423-2424, 'Z', 2426-2427, 'Z', 2429-2430, 'Z', 2432-2433, 'Z', 2435-2436, 'Z', 2438-2439, 'Z', 2441-2442, 'Z', 2444-2445, 'Z', 2447-2448, 'Z', 2450-2451, 'Z', 2453-2454, 'Z', 2456-2457, 'Z', 2459-2460, 'Z', 2462-2463, 'Z', 2465-2466, 'Z', 2468-2469, 'Z', 2471-2472, 'Z', 2474-2475, 'Z', 2477-2478, 'Z', 2480-2481, 'Z', 2483-2484, 'Z', 2486-2487, 'Z', 2489-2490, 'Z', 2492-2493, 'Z', 2495-2496, 'Z', 2498-2499, 'Z', 2501-2502, 'Z', 2504-2505, 'Z', 2507-2508, 'Z', 2510-2511, 'Z', 2513-2514, 'Z', 2516-2517, 'Z', 2519-2520, 'Z', 2522-2523, 'Z', 2524-2525, 'Z', 2527-2528, 'Z', 2530-2531, 'Z', 2533-2534, 'Z', 2536-2537, 'Z', 2539-2540, 'Z', 2542-2543, 'Z', 2545-2546, 'Z', 2548-2549, 'Z', 2551-2552, 'Z', 2554-2555, 'Z', 2557-2558, 'Z', 2560-2561, 'Z', 2563-2564, 'Z', 2566-2567, 'Z', 2569-2570, 'Z', 2572-2573, 'Z', 2575-2576, 'Z', 2578-2579, 'Z', 2581-2582, 'Z', 2584-2585, 'Z', 2587-2588, 'Z', 2590-2591, 'Z', 2593-2594, 'Z', 2596-2597, 'Z', 2599-2600, 'Z', 2602-2603, 'Z', 2604-2605, 'Z', 2607-2608, 'Z', 2610-2611, 'Z', 2613-2614, 'Z', 2616-2617, 'Z', 2619-2620, 'Z', 2622-2623, 'Z', 2624-2625, 'Z', 2627-2628, 'Z', 2630-2631, 'Z', 2633-2634, 'Z', 2636-2637, 'Z', 2639-2640, 'Z', 2642-2643, 'Z', 2645-2646, 'Z', 2648-2649, 'Z', 2651-2652, 'Z', 2654-2655, 'Z', 2657-2658, 'Z', 2660-2661, 'Z', 2663-2664, 'Z', 2666-2667, 'Z', 2669-2670, 'Z', 2672-2673, 'Z', 2675-2676, 'Z', 2678-2679, 'Z', 2681-2682, 'Z', 2684-2685, 'Z', 2687-2688, 'Z', 2690-2691, 'Z', 2693-2694, 'Z', 2696-2697, 'Z', 2699-2700, 'Z', 2702-2703, 'Z', 2705-2706, 'Z', 2708-2709, 'Z', 2711-2712, 'Z', 2714-2715, 'Z', 2717-2718, 'Z', 2720-2721, 'Z', 2723-2724, 'Z', 2726-2727, 'Z', 2729-2730, 'Z', 2732-2733, 'Z', 2735-2736, 'Z', 2738-2739, 'Z', 2741-2742, 'Z', 2744-2745, 'Z', 2747-2748, 'Z', 2750-2751, 'Z', 2753-2754, 'Z', 2756-2757, 'Z', 2759-2760, 'Z

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1187-1194, 'C', 1196-1220 <COH>

A:Cross-references: GB:M23213; NID:G340842; PIDN:AAB59363.1; PID:G499622

A>Note: mutant sequence from a patient with mild osteogenesis imperfecta

R:Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.

Nucleic Acids Res. 16, 349, 1988

A>Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 97.0%; Score 1149; DB 1; Length 1464;

Best Local Similarity 93.6%; Pred. No. 1.8e-56;

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGCGSGRGGPGGAGVAGPKPAGERGSPGPA 48

DB 458 GPPGPAGGEGKRGARGEPTGLPGPPGCGSGRGGPGGAGVAGPKPAGERGSPGPA 517

QY 49 GPKGSPGAGRPGGAGLPGKGLTSGSPGPGDGTGTPGAGDGRGPPGPPGARGQA 108

DB 518 GPKGSPGAGRPGGAGLPGKGLTSGSPGPGDGTGTPGAGDGRGPPGPPGARGQA 577

QY 109 GVMGPPGPKGAAGEGPKAGRGVPPGPGAVGPGAGKDGAGAGQPPGAPAGERGEOGPA 168

DB 578 GVMGPPGPKGAAGEGPKAGRGVPPGPGAVGPGAGKDGAGAGQPPGAPAGERGEOGPA 637

QY 169 GSPGFGQLPGPAGPPGEGAGKPGEGQVPGDLAGPSPGAG 208

DB 638 GSPGFGQLPGPAGPPGEGAGKPGEGQVPGDLAGPSPGAG 677

RESULT 2

CURTIS

collagen alpha 1(I) chain - rat (tentative sequence) (fragments)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 13-Jul-1981 #sequence revision 13-Jul-1981 #test change 31-Mar-2000

C:Accession: A90559; A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209; A91

R:Bornstein, P. 63-71, 1969

A>Title: Comparative sequence studies of rat skin and tendon collagen. II. The absence o

A:Reference number: A90559; MUID:69155173; PMID:5777344

A:Contents: CNB0 and CNB1

A:Accession: A90559

A:Molecule type: protein

A:Residues: 1-19 <B01>

A:Experimental source: tendon

A>Note: sequences from skin and tendon appear to be identical

A>Note: the amino-terminal tetrapeptide may be removed by limited proteolysis during ext

R:Kang, A.H.; Bornstein, P.; Piez, K.A.

Biochemistry 6, 788-795, 1967

A>Title: The amino acid sequence of peptides from the cross-linking region of rat skin c

A:Reference number: A90552; MUID:67162268; PMID:5337886

A:Contents: CNB1

A:Accession: A90552

A:Molecule type: protein

A:Residues: 5-19 <KAN>

A:Experimental source: skin

R:Bornstein, P.

J. Biol. Chem. 242, 2572-2574, 1967

A>Title: The incomplete hydroxylation of individual prolyl residues in collagen.

A:Reference number: A92029; MUID:67165368; PMID:4290711

A:Contents: CNB2

A:Accession: A92029

A:Molecule type: protein

A:Residues: 20-55 <B02>

A:Experimental source: skin and tendon

R:Butler, W.T.; Ponds, S.B.

Biochemistry 10, 2076-2081, 1971

A>Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a

A:Reference number: A90353; MUID:71263178; PMID:4327399

A:Contents: CNB4

A:Accession: A90353

A:Molecule type: protein

A:Residues: 56-102 <B01>

A:Experimental source: skin

R:Butler, W.T.

Biochemistry 9, 44-50, 1970

A>Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cove

A:Reference number: A90566; MUID:70085124; PMID:5411206

A:Contents: CNB5

A:Accession: A90566

A:Molecule type: protein

A:Residues: 103-139 <B02>

A:Experimental source: skin

R:Balian, G.; Click, E.M.; Bornstein, P.

Biochemistry 10, 4470-4478, 1971

A>Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylan

A:Reference number: A90357; MUID:72136131; PMID:4335087

A:Contents: CNB8

A:Accession: A90357

A:Molecule type: protein

A:Residues: 140-238 <B01>

A:Experimental source: skin

R:Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.

Biochemistry 11, 3798-3806, 1972

A>Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylan

A:Reference number: A90362; MUID:73006942; PMID:4342027

A:Contents: CNB8

A:Accession: A90362

A:Molecule type: protein

A:Residues: 239-418 <B02>

A:Experimental source: skin

R:Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.

Biochemistry 13, 2946-2953, 1974

A>Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino ac

A:Reference number: A90379; MUID:74271984; PMID:4366532

A:Contents: CNB3

A:Accession: A90379

A:Molecule type: protein

A:Residues: 419-567 <B03>

A:Experimental source: skin

R:Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.

Eur. J. Biochem. 37, 287-294, 1973

A>Title: Structural and immunogenic properties of a major antigenic determinant in neutra

A:Reference number: A91209; MUID:74011954; PMID:4126850

A:Contents: CNB6

A:Accession: A91209

A:Molecule type: protein

A:Residues: 568-651 <ST1>

A:Experimental source: skin

A>Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain

A>Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) in

R:Stoltz, M.; Timpl, R.; Kuehn, K.

FEBS Lett. 26, 61-65, 1972

A>Title: Non-helical regions in rat collagen alpha1-chain.

A:Reference number: A91385; MUID:73049495; PMID:4636751

A:Contents: CNB6

A:Accession: A91385

A:Molecule type: protein

A:Residues: 651-671 <ST2>

A:Experimental source: skin

A>Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequen

A>Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C

ed and subsequently O-glycosylated.

C:Comment: The order of the nine CNB peptides in the alpha 1(I) chain of rat skin colla

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl;

F:1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status

F:9/Modified site: allysine (Lys) #status experimental

F:103/424,547/Binding site: carboxylate (Lys) (covalent) #status experimental

F:103/Modified site: 5-hydroxylysine (Lys) #status experimental

F:424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 94.4%; Score 1118; DB 1; Length 671;

Best Local Similarity 90.0%; Pred. No. 4.9e-55;

Matches 198; Conservative 5; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPQPCGGRGSGRPGADGVAGPKGAPAGERSGCPA 48
|||
Db 296 GPPGAGBEGKRGARGEPGSLGPPGCRGGRGPGADGVAGPKGAPAGERSGCPA 355
|||
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGPGAGDGRPGPPGARGQA 108
|||
Db 356 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGPGAGZBGRPGAGPPGARGQA 415
|||
QY 109 GVMGFPFGKGAAGPGKAGRGVPGGAVGAPGAGKDGAGAGQGPAGAGERGEQCPA 168
|||
Db 416 GVMGFPFGKGTAGBFGKAGRGVPGGAVGAPGAGKDGAGAGQGPAGAGERGEQCPA 475
|||
QY 169 GSPGFQGLPGPAGPPGEGAKRGEGQVPGDLGAPGSPGAG 208
|||
Db 476 GSPGFQGLPGPAGPPGEGAKRGEGQVPGDLGAPGSPGARG 515
|||

RESULT 3

S21626

collagen alpha 1(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004

C:Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626

R:Li, S.W.; Kuillan, J.; Prockop, D.J.

Matrix Biol. 14, 593-595, 1994

A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I

A:Reference number: S57243

A:Accession: S57243

A:Molecule type: mRNA

A:Residues: 1-1453 <LIS>

A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:9470673; PIDN:AAA8912.1; PID:94706

R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16374

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1453 <MET>

A:Cross-references: EMBL:X57981; NID:950484; PIDN:CAA41046.1; PID:g50485

R:French, B.T.; Lee, W.H.; Maul, G.G.

Gene 39, 311-312, 1985

A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.

A:Reference number: A23982; MUID:86137403; PMID:3841523

A:Accession: A23982

A:Molecule type: mRNA

A:Residues: 518-1128 <PRE>

A:Cross-references: GB:M14423; NID:g192261; PIDN:AAA37333.1; PID:g192262

R:Monson, J.M.; Friedman, J.; McCarthy, B.J.

Mol. Cell. Biol. 2, 1362-1371, 1982

A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for

A:Reference number: I49559; MUID:83141374; PMID:6298597

A:Accession: I49559

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 735-1130 <RES>

A:Cross-references: GB:M17491; NID:g192263; PIDN:AAA37334.1; PID:g192264

R:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.

Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984

A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads

A:Reference number: I49557; MUID:84170331; PMID:6324198

A:Accession: I49557

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-25 <RE2>

A:Cross-references: GB:X01686; NID:g192246; PIDN:AAA37330.1; PID:g553881

R:Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.

Biochim. Biophys. Acta 1216, 469-474, 1993

A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.

A:Reference number: S39789; MUID:94092741; PMID:8268229

A:Accession: S39789

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1

R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.

Mol. Cell. Biol. 14, 5950-5960, 1994

A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect

A:Reference number: I48300; MUID:94344105; PMID:8065328

A:Accession: I48300

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

A:Cross-references: EMBL:X54876; NID:g50486; PIDN:CAA38657.1; PID:g50487

C:Genetics:

A:Gene: COL1A1

A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix

F:1-23/Domain: signal sequence #status predicted <SIG>

F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>

F:30-89/Domain: von Willebrand factor type C repeat homology <WVC>

F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>

F:1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 94.3%; Score 1116; DB 2; Length 1453;

Best Local Similarity 90.5%; Pred. No. 1.1e-54;

Matches 199; Conservative 5; Mismatches 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPQPCGGRGSGRPGADGVAGPKGAPAGERSGCPA 48
|||
Db 447 GPPGAGBEGKRGARGEPGSLGPPGCRGGRGPGADGVAGPKGAPAGERSGCPA 506
|||
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGPGAGDGRPGPPGARGQA 108
|||
Db 507 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGPGAGDGRPGPPGARGQA 566
|||
QY 109 GVMGFPFGKGAAGPGKAGRGVPGGAVGAPGAGKDGAGAGQGPAGAGERGEQCPA 168
|||
Db 567 GVMGFPFGKGTAGBFGKAGRGVPGGAVGAPGAGKDGAGAGQGPAGAGERGEQCPA 626
|||
QY 169 GSPGFQGLPGPAGPPGEGAKRGEGQVPGDLGAPGSPGAG 208
|||
Db 627 GSPGFQGLPGPAGPPGEGAKRGEGQVPGDLGAPGSPGARG 666
|||

RESULT 4

COCCHIS

collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)

C:Species: Gallus gallus (chicken)

C:Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text_change 31-Mar-2000

C:Accession: A90458; A90181; A02857

R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.

Biochemistry 21, 2048-2055, 1982

A:Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete primary

A:Reference number: A90458; MUID:82231995; PMID:7093229

A:Accession: A90458

A:Molecule type: protein

A:Residues: 1-1036 <HIG>

A:Experimental source: skin

A:Note: This is the latest in a series of papers from these workers elucidating the sequ

R:Eyre, D.R.; Glimcher, M.J.

Biochem. Biophys. Res. Commun. 48, 720-726, 1972

A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp

A:Reference number: A90181; MUID:72243016; PMID:5047697

A:Accession: A90181

A:Molecule type: protein

A:Residues: 1037-1042 <EYR>

A:Experimental source: skin

A:Note: residues 1037-1042 above correspond to the carboxyl end of the protein

C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some

C:Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pos

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone; trimet; t

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

A;Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T', 'A'; Cross-references: GB:L00977; NID:g180812; PIDN:AA23914.1; PID:g258774
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence were
A;Note: this translation is not annotated and this publication is not cited in GenBank
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184, 'GPSKDGANGIPGPI', 1185-1199, '<TII2>
A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487, '<CHE>
A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA
A;Residues: 1175-1487, '<ELI>
A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 217, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE', 1244-1246, 'N', 1248, 'X', 1250-1265, 1295-1305, 1395-1408, '<VAN>
A;Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal prope
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human type
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295, '<STR1>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909, 'PE', '<STR2>
A;Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pro
Gene.
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358, '<NUN2>
A;Cross-references: GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
R;Sangioorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28; 'R', 99-114; 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1
A;Accession: I84453
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 7-28, '<SAN2>

A;Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 541-560, '<SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A;Accession: I37251
Query Match 72.8%; Score 862; DB 1; Length 1487;
Best Local Similarity 74.0%; Pred. No. 1.1e-40;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;
QY 1 GPPGSPGPTGLPGPPGSGSGSGFGPCGADGVAGPKGAGRGSPGPGAGPKSGPGEAGRP 60
Db 482 GARGPFGVGIPGPPGEEGAGPENGFGQDGLAGPKGAPGSGSLAGPKGANGDGRP 551
QY 61 GEAGLPGAKGLTSGSPGSPGDKTGPFGAGQDGRPGPGPGARGQAGVGMFGPKGAA 120
Db 552 GEPGLPGARGLTGRPDAGPQGVKVPSPGAPGEDGRPGPGQARGQGVGMFGPKGAN 611
QY 121 GEPKAGRGVPGPCAVGAPGAKDGEAGAQPPGAPGAGRGQGPAGSPGQGLPGPA 180
Db 612 GEPKAGRGVPGPCAVGAPGAKDGEAGAQPPGAPGAGRGQGPAGSPGQGLPGPA 671
QY 181 GPPGEGAGKPGQGVPGDGLGAPGSPGAG 208
Db 672 GPPGEGAGKPGQGVPGGAGAPGLVPRG 699
RESULT 6
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N;Alternate names: type II collagen
C;Species: Equus caballus (domestic horse)
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C;Accession: T45467
R;Richardson, D.W.; Dodge, G.R.
Submitted to the EMBL Data Library, June 1996
A;Description: Cloning of equine type II collagen and modulation of its expression in eq
A;Reference number: 222977
A;Accession: T45467
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1418, '<RIC>
A;Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:AA805773.1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 72.6%; Score 860; DB 2; Length 1418;
Best Local Similarity 74.0%; Pred. No. 1.3e-40;
Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;
QY 1 GPPGSPGPTGLPGPPGSGSGSGFGPCGADGVAGPKGAGRGSPGPGAGPKSGPGEAGRP 60
Db 423 GARGPFGVGIPGPPGEEGAGPENGFGQDGLAGPKGAPGSGSLAGPKGANGDGRP 482
QY 61 GEAGLPGAKGLTSGSPGSPGDKTGPFGAGQDGRPGPGPGARGQAGVGMFGPKGAA 120
Db 483 GEPGLPGARGLTGRPDAGPQGVKVPSPGAPGEDGRPGPGQARGQGVGMFGPKGAN 542
QY 121 GEPKAGRGVPGPCAVGAPGAKDGEAGAQPPGAPGAGRGQGPAGSPGQGLPGPA 180
Db 543 GEPKAGRGVPGPCAVGAPGAKDGEAGAQPPGAPGAGRGQGPAGSPGQGLPGPA 602
QY 181 GPPGEGAGKPGQGVPGDGLGAPGSPGAG 208
Db 603 GPPGEGAGKPGQGVPGGAGAPGLVPRG 630
RESULT 7
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)

F:31-91/Domain: von Willebrand factor type C repeat homology <VMC>
F:1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 70.9%; Score 839; DB 2; Length 1487;
Best Local Similarity 72.1%; Pred. No. 1.9e-39;
Matches 150; Conservative 12; Mismatches 46; Indels 0; Gaps 0;

QY 1 GPFGPGPTGLGPPDGERGGPSRGPAGADGVAGPKPAGERGSPQAPGKSGPEAGR 60
Db | | | | | | | | | | | | | | | | | | | | | | : | | | | :
492 GARGEPGGAGTGPFGERGAPGNRGFPQODGLAGPKGAPGERGPSGLAGPKGANGDPGR 551
| | | | | | | | | | | | | | | | | | | | | | : | | | | :
QY 61 GEAGLPGAKGLTGSPGSDCKTPGPGAGODGRGPPGPPGARGOAGWGMFPDGKAA 120
Db | | | | | | | | | | | | | | | | | | | | | | : | | | | :
552 GFPLPGARGLTRGDAGPQKVPSGAPGDDGRPPGPGQARGQPQVGMFPDGKGAN 611
| | | | | | | | | | | | | | | | | | | | | | : | | | | :
QY 121 GFPKGVRGVPVGAVGPACKDGEAGAQGPPGAPAGERGEQGPAQSFGQLPGPA 180
Db | | | | | | | | | | | | | | | | | | | | | | : | | | | :
612 GFPGVAGEKGLAGAPGLRLPGKOGETGAAGPPGSPGAGERGEQAFGPSFGQLPGPP 671
| | | | | | | | | | | | | | | | | | | | | | : | | | | :
QY 181 GPPGAGKPGEGVPGDLGAPGSPGAP 208
Db | | | | | | | | | | | | | | | | | | | | | | : | | | | :
672 GPPGEGGKQGGIPGEGAGAPLVGPRG 699
| | | | | | | | | | | | | | | | | | | | | | : | | | | :

RESULT 11

CBOIS

C:Species: Bos primigenius taurus (cattle)
C:Collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)
C:Date: 24-Apr-1984 #sequence revision 31-Dec-1993 #text change 09-Jul-2004
C:Accession: A91193; A91229; A91387; A91201; A91200; A43048; A02853
R:Rauteberg, J.; Timpi, R.; Furthmayr, H.
Eur. J. Biochem. 27, 231-237, 1972
A>Title: Structural characterization of N-terminal antigenic determinants in calf and human
A:Reference number: A91193; PMID:7225334; PMID:4115172
A:Accession: A91193
A:Molecule type: protein
A:Residues: 1-19 <RAU>
A:Cross-references: UNIPROT:P02453
A:Experimental source: skin
A>Note: The epsilon carbon of Lys-9, by homology with the rat alpha 1(I) chain, is converted to proline.
R:Fietzek, P.P.; Kuehn, K.
Eur. J. Biochem. 52, 77-82, 1975
A>Title: The covalent structure of collagen: amino-acid sequence of the cyanogen-bromide treated peptide
A:Reference number: A91229; PMID:76022320; PMID:11164916
A:Accession: A91229
A:Molecule type: protein
A:Residues: 20-145 <RFI>
A:Experimental source: skin
A>Note: Lys-103 is hydroxylated and binds glucosylgalactose
R:Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.
FEBS Lett. 26, 74-76, 1972
A>Title: The covalent structure of collagen: amino acid sequence of alaphal-CB3 from calf
A:Reference number: A91387; PMID:73049499; PMID:4673951
A:Accession: A91387
A:Molecule type: protein
A:Residues: 146-294 <FT2>
A:Experimental source: skin
R:Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
Eur. J. Biochem. 38, 396-400, 1973
A>Title: The covalent structure of collagen. 2. The amino-acid sequence of alaphal-CB7 from calf
A:Reference number: A91211; PMID:74086118; PMID:4359390
A:Accession: A91211
A:Molecule type: protein
A:Residues: 295-562 <FT3>
A:Experimental source: skin
R:Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
Eur. J. Biochem. 30, 169-183, 1972
A>Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
A:Reference number: A91201; PMID:73042276; PMID:4343808
A:Accession: A91201
A:Molecule type: protein
A:Residues: 563-675 <WEN>
A:Experimental source: skin

R;Myers, J.C.; Loidl, H.R.; Stollie, C.A.; Seyer, J.M.
 J. Biol. Chem. 260, 5533-5541, 1985
 A;Title: Partial covalent structure of the human alpha 2 type V collagen chain.
 A;Reference number: I55239; MUID:85182703; PMID:298598
 A;Accession: I55239
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1002-1226 <RES>
 A;Cross-references: GB:M10956; NID:g180427; PIDN:AAA52007.1; PID:g180428
 A;Note: Part of this sequence were determined by protein sequencing
 R;Emanuel, B.S.; Cammizaro, L.A.; Seyer, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A;Reference number: I59025; MUID:85216505; PMID:3858826
 A;Accession: I59025
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1003-1034 <RES>
 A;Cross-references: GB:M1135; NID:g179693; PIDN:AAA51857.1; PID:g179694
 A;Note: Part of this sequence were determined by protein sequencing
 R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
 J. Biol. Chem. 260, 11216-11222, 1985
 A;Title: Complete primary structure of the human alpha-2 type V procollagen COOH-terminal
 A;Reference number: A25374; MUID:85289337; PMID:2411731
 A;Accession: A25374
 A;Molecule type: mRNA
 A;Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <MYE>
 A;Cross-references: GB:M11718; NID:g180912; PIDN:AAA52058.1; PID:g180913
 A;Experimental source: normal fibroblasts
 R;Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeld, C.S.; Weil, D.; Ramirez, F.
 Genomics 3, 275-277, 1988
 A;Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on
 A;Reference number: A30017; MUID:89138450; PMID:3224983
 A;Accession: A30017
 A;Molecule type: DNA
 A;Residues: 1449-1463, 'E', 1465-1495, 'A' <TSI>
 A;Cross-references: GB:J03051; NID:g179695; PIDN:AAA51859.1; PID:g179696
 A;Note: The authors translated the codon GAA for residue 1460 as Gln, and GAG for residue
 C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 are 5-hydroxylated and subsequently O-glycosylated.
 C;Comment: The amino-terminal propeptide domain appears not to be completely cleaved.
 C;Genetics:
 A;Gene: GDB:COL5A2
 A;Cross-references: GDB:119064; OMIM:120190
 A;Map position: 2q31-2q31
 A;Introns: 33/1; 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
 C;Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHUIV), a
 alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
 engh, is formed with desmosine cross-links made from lysine and allysine residues
 C;Function:
 A;Description: structural component of extracellular fibrous polymer associated with cell
 A;Note: may play a role in controlling the lateral growth of collagen I fibrils
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
 F;1-26/Domain: signal sequence #status predicted <SIG>
 F;27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>
 F;27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>
 F;40-98/Domain: nonhelical
 F;109-186/Region: von Willebrand factor type C repeat homology <VWC>
 F;187-208/Region: helical
 F;209-1225/Region: nonhelical
 F;503-505/Region: cell attachment (R-G-D) motif
 F;941-943/Region: cell attachment (R-G-D) motif
 F;1064-1066/Region: cell attachment (R-G-D) motif
 F;1067-1069/Region: cell attachment (R-G-D) motif
 F;1097-1099/Region: cell attachment (R-G-D) motif
 F;1124-1126/Region: cell attachment (R-G-D) motif
 F;1133-1135/Region: cell attachment (R-G-D) motif
 F;1225-1250/Region: carboxyl-terminal nonhelical
 F;1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>
 F;1269-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F;27/Modified site: pyrrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;193-194/Cleavage site: Ala-Gln (procollagen N-endorpeptidase) #status predicted
 F;194/Modified site: pyrrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F;201/Modified site: allysine (Lys) #status predicted
 F;290,293,296,608,614,1004,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #sta
 F;299,1139/Modified site: 5-hydroxylysine (Lys) #status predicted
 F;299,1139/Binding site: carboxylate (Lys) #status predicted
 F;1025/Modified site: 5-hydroxylysine (Lys) (covalent) #status predicted
 F;1250-1251/Cleavage site: Glu-Asp (procollagen C-endorpeptidase) #status predicted
 F;1259,1397/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;1293,1299,1325/Disulfide bonds: interchain #status predicted
 F;1333-1494,1402-1447/Disulfide bonds: #status predicted

Query Match 66.4%; Score 786; DB 1; Length 1496;
 Best Local Similarity 67.3%; Pred. No. 1.6e-36;
 Matches 140; Conservative 15; Mismatches 53; Indels 0; Gaps 0;

QY	1	GPPGPGTGLPGPPGPGSGRGGPGAGCAGVAGKGPAGRGSPGAGKSPGAGRGP	60
DB	501	GPRGPGTGLPGPPGPGVGERGAFNGFGSDGLPGFKAQGGRGVPVSGSGKSGQDFRP	560
QY	61	CEAGLPGAKGLTGTSPGSPDGKTPPGPAGQDGRPPGPPGARGQAGVGMFPKPGAA	120
DB	561	CEPLPGARGLTGNPGVQGPGRGLGLGAPGEDGRPGPGSIGIKGQPGTGLPGPKGSN	620
QY	121	CEPGKAGRGVPGPGVAGVGPAGKDEAGAGGPGGPGAGPAGRGEGQGAGSGFGLPGPA	180
DB	621	GDPGKPGVAGNPGVPGQRGAPGKDGKVGYPGPPGPGGLRGGRGQSGPPGTGFGHPGPP	680
QY	181	GPPGKAGKPGQGVPGDLGAPGSPGAG	208
DB	681	GPPGEGKPGDQGVPGGPGVAGVPLGPRG	708

RESULT 15

CGHUIV
 collagen alpha 1(III) chain precursor - human
 N;Alternate names: procollagen alpha 1(III) chain
 C;Species: Homo sapiens (man)
 C;Date: 24-Apr-1994 #sequence revision 01-Sep-1995 #text change 09-Jul-2004
 A;Accession: S05272; S04642; F80011; S01726; S04887; A90399; A94562; I51868; S59511; A904
 R;Prockop, D.J.
 submitted to the EMBL Data Library, February 1989
 A;Reference number: S05272
 A;Accession: S05272
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-1240, 'V', 1242-1466 <PRC>
 A;Cross-references: UNIPROT:P02461; EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
 R;Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
 Biochem. J. 260, 509-516, 1989
 A;Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human
 erences.
 A;Reference number: S04642; MUID:89350838; PMID:2764886
 A;Accession: S04642
 A;Molecule type: mRNA
 A;Residues: 1-1196 <ALA>
 A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
 A;Note: the complete sequence is not shown
 R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
 Gene 78, 255-265, 1989
 A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene ((
 A;Reference number: F80011; MUID:89378752; PMID:2777083
 A;Accession: F80011
 A;Molecule type: DNA
 A;Residues: 1-176 <BEN>
 A;Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814
 R;Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
 Nucleic Acids Res. 16, 7201, 1988
 A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pref
 A;Reference number: S01726; MUID:89303360; PMID:3405773
 A;Accession: S01726
 A;Molecule type: mRNA
 A;Residues: 1-170 <TOM>

A;Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061
A;Note: the authors translated the codon CAG for residue 154 as His
R;Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A;Reference number: S04887; MUID:89386015; PMID:2780304
A;Accession: S04887
A;Molecule type: mRNA
A;Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634.
A;Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R;Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A;Reference number: A90399; MUID:77134724; PMID:557335
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEV1>
A;Experimental source: liver
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose
R;Seyer, J.M.
submitted to the Atlas, December 1977
A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEV2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multigene deletion with unusual
ispring.
A;Reference number: I51868; MUID:93304430; PMID:8317500
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 186-194 <ML>
A;Cross-references: GB:S62925; NID:G386425; PIDN:AA013937.1; PID:G4261637
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1
A;Reference number: S95511; MUID:96067614; PMID:7487954
A;Accession: S95511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: GB:S79877; NID:G1195576; PIDN:AA035615.1; PID:G1195577
R;Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides
A;Reference number: A90414; MUID:79000343; PMID:687591
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675, 'N', 677-727 <SEV3>
A;Experimental source: liver
A;Note: sequence corrected by A90414
R;Lee, B.; Vitale, E.; Superi-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of the
A;Reference number: I55349; MUID:91161621; PMID:1672129
A;Accession: I55349
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from type
A;Reference number: A90438; MUID:80198282; PMID:6246925
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895, 'A', 897-964 <SEV4>
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lemande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990

A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and
A;Reference number: A98303; MUID:91009133; PMID:2145268
A;Accession: A98303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA059383.1; PID:G1
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome
R;Mankoo, B.S.; Dalglish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from type
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1
A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Meyer
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1
R;Miskulin, M.; Dalglish, R.; Kluge-Becker, B.; Rennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804; PMID:3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Cross-references: GB:M13146; NID:G180415; PIDN:AAA52003.1; PID:G180416
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID:G180417; PIDN:AAA52004.1; PID:G180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. P.
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB:
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given for
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O-
C;Genetics:
A;Gene: GDB:COL3A1
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Danlos
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide br
er of their length, is formed with desmosine cross-links made from lysine and allysine re
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains integ

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:17:54 ; Search time 31.0018 Seconds
(without alignments)
3452.202 Million cell updates/sec

Title: US-10-658-989A-1

Perfect score: 1184

Sequence: 1 GPPGEPGPTGLPFPGERGG.....GEQGVPGDLGAPGSPGAG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1149	97.0	1069	2	Q61AN8
2	1149	97.0	1460	1	CA11_CANFA
3	1149	97.0	1461	2	O76045
4	1149	97.0	1464	1	CA11_HUMAN
5	1149	97.0	1464	2	Q8N473
6	1122	94.8	1453	2	Q63079
7	1118	94.4	671	1	CA11_RAT
8	1116	94.3	1225	2	Q6PCL3
9	1116	94.3	1453	1	CA11_MOUSE
10	1116	94.3	1453	2	Q810J9
11	1067	90.1	1453	1	CA11_CHICK
12	989	83.5	1450	2	Q9YIB4
13	981	82.9	1445	2	Q93251
14	974.5	82.3	1449	2	Q640B2
15	962.5	81.3	1449	2	Q802B5
16	941	79.5	1447	2	Q9IB91
17	913	77.1	747	1	CA12_BOVIN
18	862	72.8	1418	1	CA12_HUMAN
19	862	72.8	1447	2	Q6UIJ5
20	860	72.6	1418	2	Q28396
21	860	72.6	1487	2	Q77753
22	859	72.6	1160	2	Q14046
23	859	72.6	1269	2	Q7T227
24	859	72.6	1420	2	Q90W37
25	859	72.6	1487	2	Q14047
26	853	72.0	1447	2	Q6P4U1
27	852	72.0	1492	2	Q6P4Z2
28	850	71.8	1486	2	Q91717
29	850	71.8	1486	2	Q7ZTI6
30	844	71.3	1491	2	Q91718
31	844	71.3	1491	2	Q7ZTM4

32	839	70.9	1419	2	Q63123
33	839	70.9	1442	2	Q62031
34	839	70.9	1442	2	Q62033
35	839	70.9	1459	1	CA12_MOUSE
36	839	70.9	1459	2	Q62032
37	835	70.5	1419	2	Q80VY3
38	835	70.5	1419	2	Q80X38
39	835	70.5	1487	2	Q641K3
40	833	70.4	1418	2	Q9W7K9
41	833	70.4	1449	2	Q910C0
42	824	69.6	779	1	CA11_BOVIN
43	800	67.6	1449	2	Q6NZ15
44	800	67.6	1449	2	Q6PEI9
45	798	67.4	1347	2	Q96QB3

ALIGNMENTS

RESULT 1

ID	Q61AN8	PRELIMINARY;	PRT;	1069	AA.
AC	Q61AN8;				
DT	05-JUL-2004	(TrEMBLrel. 27, Created)			
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)			
DE	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)			
DE	Collagen type I alpha 1 (fragment).				
GN	Name=COL1A1;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RM	MEDLINE=97141927; PubMed=8989177;				
RA	Simon M., Pedutour F., O'Brien K., Sirvent N., Kedra D., Guilbaud C.,				
RA	Turc-Carel C., Dumanski J.P.;				
RT	"Regulation of the platelet-derived growth factor B-chain gene via				
RT	fusion with collagen gene COL1A1 in dermatofibroblastoma.";				
RT	and giant-cell fibroblastoma.";				
RL	Nat. Genet. 15:95-98(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	O'Brien K.P.;				
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.				
EMBL	X88705; CAA67261.1; -.				
GO	GO:0005737; C:cytoplasm; IEA.				
GO	GO:0006817; P:phosphate transport; IEA.				
DR	InterPro; IPR008161; Clg_helix.				
DR	InterPro; IPR008160; Collagen.				
DR	InterPro; IPR009041; PMP_SGCI.				
DR	InterPro; IPR001007; VWF_C.				
DR	Pfam; PF01391; Collagen_16.				
DR	Pfam; PF00093; VWC_1.				
DR	ProDom; PD000007; Clg_helix; 3.				
DR	SMART; SM00214; VWC_1.				
DR	PROSITE; PS01208; VWF_C_1.				
DR	PROSITE; PS01084; VWF_C_2; 1.				
KW	Collagen.				
FT	NON_TER.				
SQ	SEQUENCE 1069 AA; 97445 MW; EE279B10572FB980 CRC64;				

Query Match 97.0%; Score 1149; DB 2; Length 1069;

Best Local Similarity 93.6%; Pred. No. 6.3e-44;

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Oy 1 GPP-----GEPGPTGLPFPGERGGSPGADGVAGPKGAGERGSPGA 48

Db 458 GPPGAGEGKRGARGEGPTGLPFPGERGGSPGADGVAGPKGAGERGSPGA 517

Oy 49 GPKSGPEAGRPAGLPGAKLTGSPGSPGDKTGPAGODGRPPGPPGARGQA 108

Db 518 GPKSGPEAGRPAGLPGAKLTGSPGSPGDKTGPAGODGRPPGPPGARGQA 577

Qy	109	GVMGFPKGAAGEPGKAGRGVPPGAVGAPGAKUGEAQAQQPAGPAGERGEQGPA	167
Dd	578	GVMGFPKGAAGEPGKAGRGVPPGAVGAPGAKUGEAQAQQPAGPAGERGEQGPA	637
Qy	169	GSPGFGLGPAGPCGACKEGVCGDVGARFSGPAG	208
Dd	638	GSPGFGLGPAGPCGACKEGVCGDVGARFSGARG	677

RESULT 2

CALL	CANFA	STANDARD;	PRT;	1460 AA.
ID	CALL CANFA			
AC	OXSJ7;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DE	25-OCT-2004 (Rel. 45, Last annotation update)			
DB	Collagen alpha 1(I) chain precursor.			
GN	Names=COL1A1;			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Cranita; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_Taxid=9615;			
[i]	SEQUENCE FROM N.A., AND VARIANT OF ALA-208.			
RN	TISSUE=Skin;			
RC	MEDLINE=21023337; PubMed=11147834; DOI=10.1006/abbi.2000.2099;			
RX	Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;			
RA	"Sequence of normal canine COL1A1 cDNA and identification of a			
RT	heterozygous alpha 1(I) collagen Gly208Ala mutation in a severe case of			
RT	canine osteogenesis imperfecta."			
RL	Arch. Biochem. Biophys. 384:37-46(2000).			
CC	-!- FUNCTION: Type I collagen is a member of group I collagen			
CC	(fibrillar forming collagen). CC			
CC	-!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.			
CC	-!- PTM: Prolines at the third position of the tripeptide repeating			
CC	unit (G-X-Y) are hydroxylated in some or all of the chains.			
CC	-!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta			
CC	(OI). CC			
CC	-!- SIMILARITY: Belongs to the fibrillar collagen family.			
CC	-!- SIMILARITY: Contains 1 VWFC domain.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (<http://www.isb-sib.ch/announcement/>) or send an email to license@lsb-sib.ch.

CC	EMBL; AF153062; AAD34619.1; "	
DR	InterPro; IPRO08161; Clg helix.	
DR	InterPro; IPRO08160; Collagen.	
DR	InterPro; IPRO00885; Fib collagen_C.	
DR	InterPro; IPRO01007; VWFC.	
DR	Pfam; PF01410; COLFI; 1.	
DR	Pfam; PF01391; Collagen; 18.	
DR	ProDom; PD000007; Clg helix; 2.	
DR	ProDom; PD002078; Fib collagen_C; 1.	
DR	PROSITE; PS01208; VWFC 1; 1.	
DR	PROSITE; PS50184; VWFC 2; 1.	
DR	Collagen; Disease mutation; Extracellular matrix; Glycoprotein;	
KW	Hydroxylation; Pyrrrolidone carboxylic acid; Repeat; Signal;	
KW	Structural protein.	
FT	SIGNAL 1 22 By similarity.	
FT	PROPEP 23 157 N-terminal propeptide.	
FT	CHAIN 158 1214 Collagen alpha 1(I) chain.	
FT	PROPEP 1215 1460 C-terminal propeptide.	
FT	DOMAIN 324 92 VWFC.	
FT	DOMAIN 158 174 Nonhelical region (N-terminal).	
FT	DOMAIN 175 1188 Triple-helical region.	
FT	DOMAIN 1189 1214 Nonhelical region (C-terminal).	
FT	SITE 741 743 Cell attachment site (Potential).	

FEBS Lett. 279:9-13 (1991).
 [5] SEQUENCE FROM N.A.
 RN MEDLINE=92157916; PubMed=1787829;
 RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
 RA Olsen A.S., Prockop D.J.;
 RT "Completion of the last half of the structure of the human gene for
 RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
 RL Matrix 11:375-379 (1991).
 [6] SEQUENCE FROM N.A.
 RN MEDLINE=98107942; PubMed=9443882;
 RA Korkko J., Ala-Korkko L., De Paepe A., Nuytinck L., Earley J.,
 RA Prockop D.J.;
 RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
 RT scanning by conformation-sensitive gel electrophoresis identifies only
 RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
 RT identification of common sequences of null-allele mutations.";
 RL Am. J. Hum. Genet. 62:98-110 (1998).
 [7] SEQUENCE FROM N.A.
 RN MEDLINE=98107942; PubMed=9443882;
 RA Korkko J., Earley J.J., Nuytinck L., DePaepe A., Prockop D.J.,
 RA Ala-Korkko L.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF017178; AAB94054.2;
 DR GO: GO:0005581; C:collagen; IEA.
 DR GO: GO:0005737; C:cytoplasm; IEA.
 DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO: GO:0006817; P:phosphate transport; IEA.
 DR InterPro: IPR008161; Clq helix.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR000885; Fib_collagen_C.
 DR InterPro: IPR009041; PMP_SGCI.
 DR InterPro: IPR001007; VWF C.
 DR Pfam: PF01410; COLFI; 1.
 DR Pfam: PF01391; Collagen; 18.
 DR ProDom: PD000007; Clq_helix; 3.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00039; COLFI; 1.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWF_C_1; 1.
 DR PROSITE: PS01184; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA7821 CRC64;

Query Match 97.0%; Score 1149; DB 2; Length 1461;
 Best Local Similarity 93.6%; Pred. No. 7, 7e-44;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GRFGPTGLPPGGERGSGRPPGADGVAGPKGAPGERGSPGPA 48
 DB |||||
 455 GPPGPGAGEGKRGARGEPTGLPPGGERGSGRPPGADGVAGPKGAPGERGSPGPA 514
 QY 49 GPKSGPGAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGODGPPGPPGARGQA 108
 DB |||||
 515 GPKSGPGAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGODGPPGPPGARGQA 574
 QY 109 GVMGPPGKGAAGEPKAGRGVPPGAVGAGKDGAGAGQPPGAPGAGERGQCPA 168
 DB |||||
 575 GVMGPPGKGAAGEPKAGRGVPPGAVGAGKDGAGAGQPPGAPGAGERGQCPA 634
 QY 169 GSPGFGQLGPPAGPPGEGAGKPGEGQGVGDLAGPPGSPGAG 208
 DB |||||
 635 GSPGFGQLGPPAGPPGEGAGKPGEGQGVGDLAGPPGSPGAG 674

RESULT 4
 CAIL HUMAN
 ID CAIL HUMAN STANDARD; PRT: 1464 AA.
 AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
 AC Q15201; Q16050; Q7K230; Q7K234; Q8IV15; Q9UML6; Q9UML7;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)

25-JAN-2005 (Rel. 46, Last annotation update)
 DT Collagen alpha 1(I) chain precursor.
 DE Name=COL1A1;
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A., AND VARIANT SER-1434.
 RA Dalglish R.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 [2] SEQUENCE OF 1-589 FROM N.A.
 RP MEDLINE=98329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
 RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
 RA Pretorius P.J.;
 RT "Complete nucleotide sequence of the region encompassing the first
 RT twenty-five exons of the human pro alpha 1(I) collagen gene
 RT (COL1A1).";
 RL Gene 67:105-115 (1988).
 [3] SEQUENCE OF 1-472 FROM N.A.
 RP MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
 RA Jaenisch R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
 RT of human type I procollagen.";
 RL Biochem. J. 253:919-922 (1988).
 [4] SEQUENCE OF 1-181 FROM N.A.
 RP MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 RA Myers J., Williams C., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons.";
 RL Nature 310:337-340 (1984).
 [5] SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
 RP TISSUE=Skin;
 RX MEDLINE=71038625; PubMed=5529814;
 RA Click E.M., Bornstein P.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706 (1970).
 [6] SEQUENCE OF 425-1464 FROM N.A.
 RP MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223 (1983).
 [7] SEQUENCE OF 472-607 FROM N.A.
 RP PubMed=2981843;
 RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
 RT "Multixon deletion in an osteogenesis imperfecta variant with
 RT increased type III collagen mRNA.";
 RL J. Biol. Chem. 260:691-694 (1985).
 [8] SEQUENCE OF 488-625 FROM N.A.
 RP PubMed=3857621;
 RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gellinas R.E.;
 RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
 RT I collagen chain in a lethal form of osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874 (1985).
 [9] SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-II HIS-1277; ARG-1388 AND
 RP 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
 RX MEDLINE=93352646; PubMed=8349697;
 RA Chessier S.D., Wallis G.A., Byers P.H.;
 RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)

RT chain of type I collagen result in defective chain association and
 RT produce lethal osteogenesis imperfecta.";
 RL J. Biol. Chem. 268:18218-18225(1993).
 RN [10]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 RC TISSUE=Bone; PubMed=3340531;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [11]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1(I) collagen gene.
 RT Promoter structure, Alu repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [12]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to transcriptional
 RT control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [13]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rosouw C.M.S., Verger W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [14]
 RP SEQUENCE OF 33-52.
 RX PubMed=2318855;
 RA Wirtz M.K., Keene D.R., Hori H., Glatville R.W., Steinmann B.,
 RA Rao V.H., Hollister D.W.;
 RT "In vivo and in vitro noncovalent association of excised alpha 1(I)
 RT amino-terminal propeptides with mutant alpha 2(I) collagen chains
 RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
 RT VII.";
 RL J. Biol. Chem. 265:6312-6317(1990).
 RN [15]
 RP SEQUENCE OF 156-183 FROM N.A.
 RX PubMed=2767050;
 RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,
 RA Bateman J.F.;
 RT "A base substitution in the exon of a collagen gene causes alternative
 RT splicing and generates a structurally abnormal polypeptide in a
 RT patient with Ehlers-Danlos syndrome type VII.";
 RL EMBO J. 8:1705-1710(1989).
 RN [16]
 RP SEQUENCE OF 175-187 AND 274-289.
 RX PubMed=2169412;
 RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
 RA Muller P.K.;
 RT "A critical crosslink region in human-bone-derived collagen type I.
 RT Specific cleavage site at residue Leu95.";
 RL Eur. J. Biochem. 192:153-159(1990).
 RN [17]
 RP SEQUENCE OF 263-268.
 RC TISSUE=Skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected vertebrate
 RT collagens. A possible role of the carbohydrate in fibril formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [18]
 RP SEQUENCE OF 281-302; 402-420; 823-842; 1026-1045 AND
 RP 1143-1162 FROM N.A.
 RX PubMed=2374517;
 RA Labhard M.E., Hollister D.W.;

RT "Segmental amplification of the entire helical and telopeptide regions
 RT of the cDNA for human alpha 1(I) collagen.";
 RL Matrix 10:124-130(1990).
 RN [19]
 RP SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
 RX MEDLINE=83064528; PubMed=6183642;
 RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
 RT "Cloning and characterization of five overlapping cDNAs specific for
 RT the human pro alpha 1(I) collagen chain.";
 RL Nucleic Acids Res. 10:5925-5934(1982).
 RN [20]
 RP SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
 RX PubMed=2339700;
 RA Wallis G.A., Starman B.J., Zimm A.B., Byers P.H.;
 RT "Variable expression of osteogenesis imperfecta in a nuclear family is
 RT explained by somatic mosaicism for a lethal point mutation in the
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
 RL Am. J. Hum. Genet. 46:1034-1040(1990).
 RN [21]
 RP SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
 RX MEDLINE=95187161; PubMed=7881420;
 RA Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
 RA Brunelli P.C., Mottes M.;
 RT "Severe (type III) osteogenesis imperfecta due to glycine
 RT substitutions in the central domain of the collagen triple helix.";
 RL Hum. Mol. Genet. 3:2201-2206(1994).
 RN [22]
 RP SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.
 RX PubMed=3170557;
 RA Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andreassen P.,
 RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
 RT "Substitution of cysteine for glycine within the carboxyl-terminal
 RT telopeptide of the alpha 1 chain of type I collagen produces mild
 RT osteogenesis imperfecta.";
 RL J. Biol. Chem. 263:14605-14607(1988).
 RN [23]
 RP SEQUENCE OF 1440-1464 FROM N.A.
 RX MEDLINE=90110490; PubMed=2295701;
 RA Willing M.C., Cohn D.H., Byers P.H.;
 RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I
 RT collagen predicts an elongated pro alpha 1(I) chain and results in
 RT osteogenesis imperfecta type I.";
 RL J. Clin. Invest. 85:282-290(1990).
 RN [24]
 RP SEQUENCE OF 1454-1464 FROM N.A.
 RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
 RA Maatta A., Bornstein P., Penttinen R.P.;
 RT "Highly conserved sequences in the 3'-untranslated region of the
 RT COL1A1 gene bind cell-specific nuclear proteins.";
 Query Match 97.0%; Score 1149; DB 1; Length 1464;
 Best Local Similarity 93.6%; Pred. No 7,7e-44;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGPTGLPGPGRGGPGSRGFGADGADGAGKPGAGERSGPGA 48
 Db |||||
 458 GPPGPAAGEGKRGARGEPGPTGLPGPGRGGPGSRGFGADGADGAGKPGAGERSGPGA 517
 QY 49 GPKGSPCEAGRPGCEAGLPGCAKGLTGSPGSPGDKGTGPPGAGODRPPGPPGARGQA 108
 Db |||||
 518 GPKGSPCEAGRPGCEAGLPGCAKGLTGSPGSPGDKGTGPPGAGODRPPGPPGARGQA 577
 QY 109 GVGWFFGPKGAAGEFGKAGERGVPPGPGAVGAGKDGAGAGQPPGPGAGPAGERGQPGA 168
 Db 578 GVGWFFGPKGAAGEFGKAGERGVPPGPGAVGAGKDGAGAGQPPGPGAGPAGERGQPGA 637
 QY 169 GSPGFGGLPGPAGPCEACKPGCEQGVPGDLGAPGSPGAG 208
 Db 638 GSPGFGGLPGPAGPCEACKPGCEQGVPGDLGAPGSPGAG 677
 RESULT 5
 Q8N473

Qy	109	GVNGFFGPGKGAAGPCKAGRGVPGPGCAVGPAGKDGCAAGCAQGGPPGAGRGEGGPA	168
Db	578	GVNGFFGPGKGAAGPCKAGRGVPGPGCAVGPAGKDGCAAGCAQGGPPGAGRGEGGPA	637
Qy	169	GSPGFQGLPGPAGPGCAAGPKGEGQGVPGDLGAPGPGSPAG	208
Db	638	GSPGFQGLPGPAGPGCAAGPKGEGQGVPGDLGAPGPGSARG	677
RESULT 6			
ID	Q63079	PRELIMINARY; PRT; 1453 AA.	
AC	Q63079;		
DT	01-JUN-1996 (TrEMBLrel. 01, Created)		
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Collagen alpha1 (Fragment).		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RI	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;		
RC	MEDLINE=99163824; PubMed=10065941;		
RA	Brandsten C., Lundmark C., Christerson C., Hammarstrom L., Wurtz T.;		
RA	"Expression of collagen alpha1(i) mRNA variants during tooth and bone		
RL	formation in the rat.;"		
RL	J. Dent. Res. 78:11-19(1999).		
DR	EMBL; Z78279; CAB01633.1; --		
DR	GO; GO:0005581; C:collagen; IEA.		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.		
DR	GO; GO:0006817; P:phosphate transport; IEA.		
DR	InterPro; IPR008161; Clg helix.		
DR	InterPro; IPR008160; Collagen.		
DR	InterPro; IPR000885; Fib_collagen_C.		
DR	InterPro; IPR009041; PMP_SGCI.		
DR	InterPro; IPR001007; VWF_C.		
DR	Pfam; PF01410; COLFI; 1.		
DR	Pfam; PF01391; Collagen; 18.		
DR	ProDom; PD000007; Clg_helix; 3.		
DR	ProDom; PD002078; Fib_collagen_C; 1.		
DR	SMART; SM00038; COLFI; 1.		
DR	SMART; SM00214; VWC; 1.		
DR	PROSITE; PS01208; VWF_C; 1.		
DR	PROSITE; PS00184; VWF_C2; 1.		
KW	Collagen.		
FT	NON TER		
Qy	SEQUENCE 1 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;		
Query Match	94.8%; Score 1122; DB 2; Length 1453;		
Best Local Similarity	91.4%; Pred. No. 1.2e-42;		
Matches	201; Conservative 2; Mismatches 5; Indels 12; Gaps 1		
Qy	1 GPP-----GEPGPTGLPGPPGERGGGSGRFFPGADGVAGPKGPAGERGSPGPA	48	
Db	447 GPGPGAGEEKGARGEPFGSGLPGPPGERGGGSGRFFPGADGVAGPKGPAGERGSPGPA	506	
Qy	49 GPKGSPGEACRGCEAGLPGAKGLTGTSPGSPGPGDKTTPPGAGQDGRPGPPGARGQA	108	
Db	507 GPKGSPGEACRGCEAGLPGAKGLTGTSPGSPGPGDKTTPPGAGQDGRPGPPGARGQA	566	
Qy	109 GVNGFFGPGKGAAGPCKAGRGVPGPGCAVGPAGKDGCAAGCAQGGPPGAGRGEGGPA	168	
Db	567 GVNGFFGPGKGTAGEPGKAGRGVPGPGCAVGPAGKDGCAAGCAQGGPPGAGRGEGGPA	626	
Qy	169 GSPGFQGLPGPAGPGCAAGPKGEGQGVPGDLGAPGPGSPAG	208	
Db	627 GSPGFQGLPGPAGPGCAAGPKGEGQGVPGDLGAPGPGSARG	666	
RESULT 7			

CALL_RAT	ID	CALL_RAT	STANDARD;	PRT;	671 AA.
AC	P02454;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	25-OCT-2004 (Rel. 45, Last annotation update)				
DE	Collagen alpha 1(I) chain (Fragments).				
GN	Name=Collal;				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
NCBI_TaxID=10116;					
RP	SEQUENCE OF 1-19.				
RP	MEDLINE=69155173; PubMed=5777344;				
RA	Bornstein P.;				
RT	"Comparative sequence studies of rat skin and tendon collagen. II. The				
RT	absence of a short sequence at the amino terminus of the skin alpha-1				
RT	chain.";				
RL	Biochemistry 8:63-71(1969).				
RL	[2]				
RP	SEQUENCE OF 5-19.				
RP	MEDLINE=67162268; PubMed=5337886;				
RA	Kang A.H., Bornstein P., Piez K.A.;				
RT	"The amino acid sequence of peptides from the cross-linking region of				
RT	rat skin collagen.";				
RL	Biochemistry 6:788-795(1967).				
RL	[3]				
RP	SEQUENCE OF 20-55.				
RP	MEDLINE=67165368; PubMed=4290711;				
RA	Bornstein P.;				
RT	"The incomplete hydroxylation of individual prolyl residues in				
RT	collagen.";				
RL	J. Biol. Chem. 242:2572-2574(1967).				
RL	[4]				
RP	SEQUENCE OF 56-102.				
RP	MEDLINE=71263178; PubMed=4327399;				
RA	Butler W.T., Ponds S.L.;				
RT	"Chemical studies on the cyanogen bromide peptides of rat skin				
RT	collagen. Amino acid sequence of alpha 1-CB4.";				
RL	Biochemistry 10:2076-2081(1971).				
RL	[5]				
RP	SEQUENCE OF 103-139.				
RP	MEDLINE=70085124; PubMed=5411206;				
RA	Butler W.T.;				
RT	"Chemical studies on the cyanogen bromide peptides of rat skin				
RT	collagen. The covalent structure of alpha 1-CB5, the major hexose-				
RT	containing cyanogen bromide peptide of alpha 1.;"				
RL	Biochemistry 9:44-50(1970).				
RL	[6]				
RP	SEQUENCE OF 140-238.				
RP	MEDLINE=72136131; PubMed=4335087;				
RA	Balian G., Click E.M., Bornstein P.;				
RT	"Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of				
RT	the hydroxylamine-produced fragment HA1.";				
RL	Biochemistry 10:4470-4478(1971).				
RL	[7]				
RP	SEQUENCE OF 239-418.				
RP	MEDLINE=73006942; PubMed=4342027;				
RA	Balian G., Click E.M., Hermanson M.A., Bornstein P.;				
RT	"Structure of rat skin collagen alpha 1-CBB. Amino acid sequence of				
RT	the hydroxyl amine-produced fragment HA2.";				
RL	Biochemistry 11:3798-3806(1972).				
RL	[8]				
RP	SEQUENCE OF 419-567.				
RP	MEDLINE=74271984; PubMed=4366532;				
RA	Butler W.T., Underwood S.P., Finch J.E. Jr.;				
RT	"Chemical studies on the cyanogen bromide peptides of rat skin				
RT	collagen. Amino acid sequence of alpha 1-CB3.";				
RL	Biochemistry 13:2946-2953(1974).				
RL	[9]				
RP	SEQUENCE OF 568-651.				
RP	MEDLINE=74011954; PubMed=4126850;				
RP	[10]				

Db 296 GPPGAGBEGKRGARPCSLGPPGRCGPGSRGPGGADGVAGPKGPGAGSGPCGA 355
QY 49 GPKGSPGAGRGAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 108
Db 356 GPKGSPGAGRGAGLPGCAKGLTSGSPGPDGKTGPPGAGZBGRGPPGPPGARGQA 415
QY 109 GVMGPPGKGAAGEPGKAGRGVPCGAVGPGAGKDGAGAGQDPPGPPGAGRGEGQGA 168
Db 416 GVMGPPGKGTAGEPGKAGRGVPCGAVGPGAGKDGAGAGQDPPGPPGAGRGEGQGA 475
QY 169 GSPGFGQLGPPGAPGPGGAGKPGGQGVPGDLGAPGSPGAG 208
Db 476 GSPGFGQLGPPGAPGPGGAGKPGGQGVPGDLGAPGSPGAG 515
RESULT 8
Q6PCL3
ID Q6PCL3 PRELIMINARY; PRT; 1225 AA.
AC Q6PCL3;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Collal protein.
GN Name=Collal;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC059281; AAHS9281.1; .
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 13.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 1.

KW Collagen. 1225 AA; 117860 MW; B6B86CBB4457F4D9 CRC64;
SQ SEQUENCE
Query Match 94.3%; Score 1116; DB 2; Length 1225;
Best Local Similarity 90.5%; Pred. No. 2e-42;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;
QY 1 GPP-----GSPGPTGLRPPGRCGPGSRGPGGADGVAGPKGPGAGSGSPCA 48
Db 447 GPPGAGBEGKRGARPCSLGPPGRCGPGSRGPGGADGVAGPKGPGAGSGPCGA 506
QY 49 GPKGSPGAGRGAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 108
Db 507 GPKGSPGAGRGAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 566
QY 109 GVMGPPGKGAAGEPGKAGRGVPCGAVGPGAGKDGAGAGQDPPGPPGAGRGEGQGA 168
Db 567 GVMGPPGKGTAGEPGKAGRGVPCGAVGPGAGKDGAGAGQDPPGPPGAGRGEGQGA 626
QY 169 GSPGFGQLGPPGAPGPGGAGKPGGQGVPGDLGAPGSPGAG 208
Db 627 GSPGFGQLGPPGAPGPGGAGKPGGQGVPGDLGAPGSPGAG 666
RESULT 9
CAIL_MOUSE
ID CAIL_MOUSE STANDARD; PRT; 1453 AA.
AC P11087; Q60635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=Collal; Synonyms=Collal;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen.";
RL Matrix Biol. 14:593-595 (1995).
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=85137403; PubMed=3841523; DOI=10.1016/0378-1119(85)90329-4;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I) collagen
protein.";
RL Gene 39:311-312 (1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
evidence for a mouse B1 element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371 (1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69 (1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE=88124276; PubMed=3340560;
RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size of
the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773 (1988).

CC -1- FUNCTION: Type I collagen is a member of group I collagen
 CC (fibrillar forming collagen).
 CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
 CC bones. In bones the fibrils are mineralized with calcium
 CC hydroxyapatite.
 CC -1- PM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -1- SIMILARITY: Contains 1 VWF domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U08020; AAA89912.1; -
 CC EMBL; X15896; CAA33904.1; -
 CC EMBL; M14423; AAA37333.1; -
 CC EMBL; M17491; AAA37334.1; -
 CC EMBL; X06753; CAA32927.1; -
 CC EMBL; K03036; AAA37332.1; -
 CC EMBL; K03029; AAA37332.1; JOINED.
 CC EMBL; K03030; AAA37332.1; JOINED.
 CC EMBL; K03031; AAA37332.1; JOINED.
 CC EMBL; K03032; AAA37332.1; JOINED.
 CC EMBL; K03033; AAA37332.1; JOINED.
 CC EMBL; K03034; AAA37332.1; JOINED.
 CC EMBL; K03035; AAA37332.1; JOINED.
 CC PIR; S57243; S21626.
 CC MGI; MGI:88467; Col1a1.
 CC InterPro; IPR008161; Clg_helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR000895; Fib_collagen_C.
 CC InterPro; IPR009041; FMP_SGCI.
 CC InterPro; IPR001007; VWF_SC1.
 CC Pfam; PF01410; COLFI; 1.
 CC Pfam; PF01391; Collagen; 18.
 CC ProDom; PD000007; Clg_helix; 3.
 CC ProDom; PD002078; Fib_collagen_C; 1.
 CC SMART; SM00038; COLFI; 1.
 CC SMART; SM00214; VMC; 1.
 CC PROSITE; PS01208; VWF1; 1.
 CC PROSITE; PS0184; VWF2; 1.
 CC Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;
 CC Pyridolone carboxylic acid; Repeat; Signal; Structural protein.
 CC SIGNAL 22
 CC PROPEP 23 151 N-terminal propeptide.
 CC CHAIN 152 1207 Collagen alpha 1(I) chain.
 CC PROPEP 1208 1453 C-terminal propeptide.
 CC DOMAIN 29 87 VWF.
 CC DOMAIN 152 167 Nonhelical region (N-terminal).
 CC DOMAIN 168 1181 Triple-helical region.
 CC DOMAIN 1182 1207 Nonhelical region (C-terminal).
 CC MOD_RES 152 152 Pyridolone carboxylic acid (By
 CC similarity).
 CC MOD_RES 160 160 Allysine (By similarity).
 CC MOD_RES 254 254 5-hydroxylysine (By similarity).
 CC MOD_RES 1153 1153 3-hydroxyproline (By similarity).
 CC CARBOHYD 56 N-linked (GlcNAc...) (Potential).
 CC CARBOHYD 254 O-linked (Gal...) (By similarity).
 CC CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
 CC SITE 734 736 Cell attachment site (Potential).
 CC SITE 1082 1084 Cell attachment site (Potential).
 CC CONFLICT 1450 1450 A -> V (in Ref 5).
 CC SEQUENCE 1453 AA; 137944 MW; 3B802E55DF81808 CRC64;
 CC
 CC Query Match 94.3%; Score 1116; DB 1; Length 1453;
 CC Best Local Similarity 90.5%; Pred. No. 2.2e-42;
 CC Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPQPPGGRGPGSGRFPFGADGVAGPKPGAGRGSPGPA 48
 DB |||||
 447 GPPGPAGEGKRGAGEGPGSGLPFPGRGPGSGRFPFGADGVAGPKPGSGRGAPGPA 506
 QY 49 GPKGSPGAGRPGEAGLPGAKGLTSPGSPGDKGTGPPGAGQDGRFPGPARGQA 108
 DB |||||
 507 GPKGSPGAGRPGEAGLPGAKGLTSPGSPGDKGTGPPGAGQDGRFPGPARGQA 566
 QY 109 GVMGPPGKAGRPGEAGLPGAKGLTSPGSPGDKGTGPPGAGQDGRFPGPARGQA 168
 DB |||||
 567 GVMGPPGKAGRPGEAGLPGAKGLTSPGSPGDKGTGPPGAGQDGRFPGPARGQA 626
 QY 169 GSPGQGLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 208
 DB |||||
 627 GSPGQGLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 666
 RESULT 10
 Q810J9 PRELIMINARY; PRT; 1453 AA.
 ID Q810J9
 AC Q810J9
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Procollagen, type I, alpha 1.
 GN Name=Col1a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.B., Feingold E.A., Grouse L.H., Berge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 FT "Generation and initial analysis of more than 15,000 full-length human
 FT and mouse cDNA sequences".
 FT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC050014; AAH50014.1; -
 DR MGI; MGI:88467; Col1a1.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000895; Fib_collagen_C.
 DR InterPro; IPR009041; FMP_SGCI.
 DR InterPro; IPR001007; VWF_SC1.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; VMC; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.


```
FT CONFLICT 1187 1187 F -> L (in Ref. 5).
FT SEQUENCE 1441 1441 Q -> H (in Ref. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Query Match
Best Local Similarity 90.1%; Score 1067; DB 1; Length 1453;
Matches 190; Conservative 4; Mismatches 14; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGRGGRGPGSRGPGAGVAGPKPGAGERSGPGPA 48
Db |||
447 GPPGAGEGKRGAGEGPGAGLPAGRGAPGSRGPGAGDGIAGPKPGPGRSGPGAV 506
QY 49 GPKGSPGAGRGAGEAGLPAGKGLTSGSPGPDGKTGTPPGAGQDGRGPPGPGARGQA 108
Db |||
507 GPKGSPGAGRGAGEAGLPAGKGLTSGSPGPDGKTGTPPGAGQDGRGPPGPGARGQA 566
QY 109 GVMGPPGPKGAAGEGPKGAGRGVPGGAVGAGKDGAGAGQPPGAGPAGERGEOGPA 168
Db |||
567 GVMGPPGPKGAAGEGPKGAGRGVPGGAVGAGKDGAGAGQPPGAGPAGERGEOGPA 626
QY 169 GSPFGQLGPGAGPGEAGKPGEGVPGDLGAPGSPGAG 208
Db |||
627 GAPGQGLGPGAGPGEAGKPGEGVPGDLGAPGSPGAG 666

RESULT 12
Q9YIB4
ID Q9YIB4 PRELIMINARY; PRT; 1450 AA.
AC Q9YIB4;
DT 01-WAY-1999 (TRENBLrel. 10, Created)
DT 01-WAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
RX DOI=10.1002/(SICI)1097-0177(199909)216:1<59:;AID-DVDY8>3.3.CO;2-2;
RA Asahina K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
RL and development of the blastema of regenerating newt limb.";
DR Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cyclopasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;

Query Match
Best Local Similarity 83.5%; Score 989; DB 2; Length 1450;
Matches 177; Conservative 7; Mismatches 24; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGRGGRGPGSRGPGAGVAGPKPGAGERSGPGPA 48
Db |||

Query Match
Best Local Similarity 82.9%; Score 981; DB 2; Length 1445;
Matches 178; Conservative 5; Mismatches 22; Indels 12; Gaps 2;

QY 1 GPP-----GEPGPTGLPGRGGRGPGSRGPGAGVAGPKPGAGERSGPGAGPK 51
Db |||
446 GPPGAGEGKRGAGEGPGSGPPGAGERGAPGSRGPGAGDGIAGPKPGPGRSGPGAV 505
QY 52 GSPGAGRGAGEAGLPAGKGLTSGSPGPDGKTGTPPGAGQDGRGPPGPGARGQA 111
Db |||
506 GSPGAGRGAGEAGLPAGKGLTSGSPGPDGKTGTPPGAGQDGRGPPGPGARGQA 565
QY 112 GFGPKGAAGEGPKGAGRGVPGGAVGAGKDGAGAGQPPGAGPAGERGEOGPGSP 171
Db |||
566 GFGPKGAAGEGPKGAGRGVPGGAVGAGKDGAGAGQPPGAGPAGERGEOGPGSP 625
```


Qy 172 GFQGLPGPAGPPGAGKPGEGVPGDLGAPGSPGAG 208
Db 626 GFQGLPGSPGAGSGKPGEGAGFQGV---GPSGPAG 659

RESULT 14

Q640B2 PRELIMINARY; PRT; 1449 AA.
ID Q640B2
AC Q640B2
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RA Klein S., Gerhard D.S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC082718; AAH82718.1; -;
KW Hypothetical protein.
SQ SEQUENCE 1449 AA; 137539 MW; 456639B1687A3B4B CRC64;

Query Match 82.3%; Score 974.5; DB 2; Length 1449;
Best Local Similarity 80.0%; Pred. No. 4.1e-36;
Matches 176; Conservative 9; Mismatches 20; Indels 15; Gaps 2;

Qy 1 GPP-----GEPGTGLPFGPGERGPGSRGPPGAGVAGKPGAGRGSPGA 48
Db 444 GPPGSGEGKRGSRGEPGAGPPGAGCAPGSRGPPGSDGASGKPGGPGVCPA 503
Qy 49 GPKSGPGEAGRPGENGLPGKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARGQA 108
Db 504 GAKSGPGEGRPGEPGLFGKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARGQS 563
Qy 109 GVMGPPGKGAAGEPGKAGRGVPPGPGAVGAPGAGKAGQPPGPPGAGRGEGQGA 168
Db 564 GVMGPPGKGAAGEPGKAGRGVPPGPGAVGLPGKDGAGAGQPPGPPGAGRGEGQGA 623
Qy 169 GSPGFQGLGPPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
Db 624 GPGPFQGLGPPAGPGEAGKPGEGQGVPGDV---GPSGPAG 660

RESULT 15

Query Match 81.3%; Score 962.5; DB 2; Length 1449;

Q802B5 PRELIMINARY; PRT; 1449 AA.
ID Q802B5
AC Q802B5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collal-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049829; AAH49829.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cyttoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg helix; 4.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C.1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_2; 1.
KW Collagen.
SQ SEQUENCE 1449 AA; 137464 MW; 4B32E20BECCE978C CRC64;

Search completed: June 17, 2005, 15:13:46
Job time : 34.0018 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 100.815 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-2
Perfect score: 3488
Sequence: 1 GPPGEPGPTGLPGRPGRGG.....GEQVPGDLGAPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3488	100.0	617	8	ADM48391 Human rec
2	3488	100.0	617	8	ADQ26217 Human gel
3	3482	99.8	821	8	ADM48392 Human rec
4	2237.5	64.1	1057	3	AAY84544 A human c
5	2237.5	64.1	1058	3	AAY84403 Amino aci
6	2237.5	64.1	1107	3	AAY84540 Amino aci
7	2237.5	64.1	1171	3	AAY84538 A chimeri
8	2237.5	64.1	1464	2	AAW68485 Human rec
9	2237.5	64.1	1464	7	ADD45059 Human Pro
10	2237.5	64.1	1464	7	ADD45055 Human Pro
11	2237.5	64.1	1464	7	ADD45051 Human Pro
12	2237.5	64.1	1464	8	ADR16800 Human col
13	2237.5	64.1	1464	8	ADR16425 Human col
14	2237.5	64.1	1464	8	ADR99144 Collagen,
15	2234	64.0	1449	4	AAE02535 Porcine a
16	2233.5	64.0	1161	7	ADG87050 Human pan
17	2233.5	64.0	1461	5	ABG93947 Human pol
18	2233.5	64.0	1464	4	AAU14136 Human nov
19	2233.5	64.0	1536	7	ADG87051 Human pan
20	2231.5	64.0	1057	3	AAY84541 Amino aci
21	2229.5	63.9	1388	3	AAY84539 Amino aci
22	2228.5	63.9	1341	2	AA71701 Collagen
23	2228.5	63.9	1341	3	AAY96122 Collagen
24	2228.5	63.9	1341	5	AAE16475 Human col
25	2228.5	63.9	1341	5	ABB80733 Collagen

26	2228.5	63.9	1341	5	ABB09625 Amino aci
27	2228.5	63.9	1341	7	ADF13075 Human col
28	2227.5	63.9	1464	4	AAB82454 Human pro
29	2227.5	63.9	1464	5	ABB90764 Human tum
30	2227.5	63.9	1464	5	ABP68610 Human pan
31	2227.5	63.9	1464	6	ABU54471 Human tum
32	2227.5	63.9	1464	6	ABR47417 Breast ca
33	2227.5	63.9	1464	6	ABR92064 Human cer
34	2227.5	63.9	1464	7	ADD14142 Human src
35	2227.5	63.9	1464	7	ADP65246 Human alp
36	2227.5	63.9	1464	8	ADQ19470 Human sof
37	2227.5	63.9	1464	8	ADQ29653 Human col
38	2225.5	63.8	1463	4	AAE02532 Bovine ol
39	2224.5	63.8	1464	7	ADE87048 Human pan
40	2222.5	63.7	1107	2	AAR89472 Collagen/
41	2222.5	63.7	1169	2	AAR89469 Collagen/
42	2222.5	63.7	1169	3	AAY84537 Amino aci
43	2222.5	63.7	1171	2	AAR89470 Collagen/
44	2222.5	63.7	1388	2	AAR89471 Collagen/
45	2222	63.7	1211	7	ADE87057 Human pan

ALIGNMENTS

RESULT 1
ADM48391
ID ADM48391 standard; protein; 617 AA.
XX AC ADM48391;
XX DT 03-JUN-2004 (first entry)
XX DE Human recombinant gelatin-like polypeptide Hu-3.
XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.
XX OS Homo sapiens.
XX FN EPI398324-A1.
XX PD 17-MAR-2004.
XX PF 11-SEP-2002; 2002EP-00078745.
XX PR 11-SEP-2002; 2002EP-00078745.
XX PA (FUJF) FUJI PHOTO FILM BV.
XX PI Bouwstra JB, Toda Y;
XX DR WPI; 2004-229415/22.
XX PT Composition useful as substitute for plasma, comprises solution of saline
XX PI and recombinant gelatin-like protein having colloid osmotic function.
XX FS Example 1; SEQ ID NO 2; 31pp; English.

The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a trimer of human recombinant gelatin-like protein Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgE antibodies. The gelatin-like proteins can be covalently attached to pharmaceutically

CC active compounds. After administration, the coupled medicament will not
 CC diffuse from the circulating blood into the interstitium. Clearance by
 CC liver and kidney will be kept to a minimum, ensuring a more constant
 CC plasma level of the medicament. Suitable medicaments include those
 CC involved in intervening blood clotting, vasodilation, function of
 CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood
 CC levels of messenger molecules such as hormones.
 XX
 SQ Sequence 617 AA;

Query Match 100.0%; Score 3488; DB 8; Length 617;
 Best Local Similarity 100.0%; Pred. No. 1.8e-207;
 Matches 617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGEGPTGLPDPGCGSGRFPDADGVAGKPGAGRGSPGAGKSGPGEAGRP 60
 Db 1 GPPGEGPTGLPDPGCGSGRFPDADGVAGKPGAGRGSPGAGKSGPGEAGRP 60
 QY 61 GEAGLPGAKGLTSGSPGDPKTPGPPAGQDGRPPGPPGARGQAGVMGFPKGA 120
 Db 61 GEAGLPGAKGLTSGSPGDPKTPGPPAGQDGRPPGPPGARGQAGVMGFPKGA 120
 QY 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 180
 Db 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 180
 QY 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 240
 Db 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 240
 QY 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 300
 Db 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 300
 QY 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 360
 Db 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 360
 QY 361 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 420
 Db 361 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 420
 QY 421 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 480
 Db 421 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 480
 QY 481 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 540
 Db 481 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 540
 QY 541 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 600
 Db 541 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 600
 QY 601 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 617
 Db 601 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 617

RESULT 2
 ADQ26217
 ID ADQ26217 standard; protein; 617 AA.
 XX
 AC ADQ26217;
 XX
 DT 23-SEP-2004 (first entry)
 XX
 DE Human gelatine-like polypeptide Hu-3.
 KW Human; Gelatine-like protein; Hu-3; microcarrier; cell culture.
 XX
 OS Homo sapiens.
 XX

PN WO2004056976-A2.
 XX 08-JUL-2004.
 XX 23-DEC-2003; 2003WO-NL000922.
 XX 23-DEC-2002; 2002EP-00080539.
 XX (FUJF) FUJI PHOTO FILM BV.
 XX Bouwstra JB, Van Es AJJ, Toda Y;
 XX WPI; 2004-507711/48.
 XX
 XX Preparing cell culture support useful for culturing anchorage dependent
 PT cells, involves coating microcarrier bead with gelatine or gelatine-like
 PT protein having specific molecular weight.
 XX
 PS Example 1; SEQ ID NO 1; 19pp; English.
 XX
 CC The present sequence is the protein sequence of human recombinant gelatin
 CC -like polypeptide Hu-3, which has a molecular weight of approximately 54
 CC kDa and which contains multiple Gly-Xaa-Yaa triplets. In an example from
 CC the invention, recombinant Hu-3 was immobilised on polystyrene beads
 CC using a heterobifunctional crosslinking agent, and used in a spinner
 CC flask cell culture. A claimed process for the preparation of a cell
 CC culture support comprises the step of coating a microcarrier bead with
 CC gelatine or gelatine-like protein having a molecular weight of about 40-
 CC 200 kDa, and optionally further comprising the step of immobilising the
 CC gelatine or gelatine-like protein on the microcarrier. In this process,
 CC more than 75% of the protein has a molecular weight of more than 95%
 CC of the gelatine or gelatine-like protein has the same molecular weight.
 CC The gelatine or gelatine-like protein is recombinantly produced to obtain
 CC a material of uniform molecular weight and to reduce the risk of
 CC contamination with prions. A claimed cell support consists of microbeads
 CC of 50-500 µm size coated with a gelatine-like protein consisting of at
 CC least 95% Gly-Xaa-Yaa triplets and containing at least 15% proline
 CC residues and less than 5% of hydroxyproline residues, with a molecular
 CC weight distribution showing a maximum between 40 and 200 kDa, at least
 CC 75% of the protein molecules having a molecular weight within 2% of the
 CC maximum. Large-scale production of expressed products can be accomplished
 CC with gelatine-coated microcarriers.
 XX
 SQ Sequence 617 AA;

Query Match 100.0%; Score 3488; DB 8; Length 617;
 Best Local Similarity 100.0%; Pred. No. 1.8e-207;
 Matches 617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGEGPTGLPDPGCGSGRFPDADGVAGKPGAGRGSPGAGKSGPGEAGRP 60
 Db 1 GPPGEGPTGLPDPGCGSGRFPDADGVAGKPGAGRGSPGAGKSGPGEAGRP 60
 QY 61 GEAGLPGAKGLTSGSPGDPKTPGPPAGQDGRPPGPPGARGQAGVMGFPKGA 120
 Db 61 GEAGLPGAKGLTSGSPGDPKTPGPPAGQDGRPPGPPGARGQAGVMGFPKGA 120
 QY 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 180
 Db 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 180
 QY 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 240
 Db 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 240
 QY 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 300
 Db 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 300
 QY 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 360
 Db 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGQAGSPGFGQLPGPA 360

QY 361 GPAGERGQPSGFGQGLPFPAGPCEACKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
Db |||||
QY 361 GPAGERGQPSGFGQGLPFPAGPCEACKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
Db |||||
QY 421 GPPGERGSGRSGFPFADGVAGPKPAGERSGPGAPKSGPAGRPGEAGLFGAKGLT 480
Db |||||
QY 481 GSPGSPGPDGKTGTPPAGQDGRPGPPGARGOAGVMGPPGPKGAAGBPKAGRGVP 540
Db |||||
QY 481 GSPGSPGPDGKTGTPPAGQDGRPGPPGARGOAGVMGPPGPKGAAGBPKAGRGVP 540
Db |||||
QY 541 GPPCAVGPAGKDGAGAGQPPGAPGAGERGEOGAPSGPFGQLPGPAGPPGEAGKPGEQ 600
Db |||||
QY 541 GPPCAVGPAGKDGAGAGQPPGAPGAGERGEOGAPSGPFGQLPGPAGPPGEAGKPGEQ 600
Db |||||
QY 601 GVPGLGAPGSPGAGG 617
Db 601 GVPGLGAPGSPGAGG 617

RESULT 3
ADM48392
ID ADM48392 standard; protein; 821 AA.
XX
AC ADM48392;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human recombinant gelatin-like polypeptide Hu-4.
XX
KW Plasma substitute; Gelatin-like protein; plasma expander; human.
XX
OS Homo sapiens.
XX
XX EPI398324-A1.
XX
PD 17-MAR-2004.
XX
PF 11-SEP-2002; 2002EP-00078745.
XX
PR 11-SEP-2002; 2002EP-00078745.
XX
PA (FUJF) FUJI PHOTO FILM BV.
XX
PI Bouwstra JB, Toda Y;
XX
DR WPI; 2004-229415/22.
XX
XX
CC The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a tetramer of human recombinant gelatin-like protein Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgE antibodies. The gelatin-like proteins can be covalently attached to pharmaceutically active compounds. After administration, the coupled medicament will not diffuse from the circulating blood into the interstitium. Clearance by liver and kidney will be kept to a minimum, ensuring a more constant plasma level of the medicament. Suitable

CC medicaments include those involved in intervening blood clotting,
CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,
CC immune responses, and blood levels of messenger molecules such as
CC hormones.
XX
SQ Sequence 821 AA;
Query Match 99.8%; Score 3482; DB 8; Length 821;
Best Local Similarity 100.0%; Pred. No. 5.2e-207;
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPPCEPPTGLPFPGERGSGRSGFPFADGVAGPKPAGERSGPGAPKSGPGEAGRP 60
Db 1 GPPCEPPTGLPFPGERGSGRSGFPFADGVAGPKPAGERSGPGAPKSGPGEAGRP 60
QY 61 GEAGLPGAKGLTSGPSGPDGKTGTPPAGQDGRPGPPGARGOAGVMGPPGPKGAA 120
Db 61 GEAGLPGAKGLTSGPSGPDGKTGTPPAGQDGRPGPPGARGOAGVMGPPGPKGAA 120
QY 121 GEPKAGERGVPGPAGVAGPAGKDGAGAGQPPGAPGAGERGEOGAPSGPFGQLPGPA 180
Db 121 GEPKAGERGVPGPAGVAGPAGKDGAGAGQPPGAPGAGERGEOGAPSGPFGQLPGPA 180
QY 181 GPPGEACKPGEQGVGCDLGAAPSGPAGPCEPTGLPGERGSGRSGFPFADGVAGPK 240
Db 181 GPPGEACKPGEQGVGCDLGAAPSGPAGPCEPTGLPGERGSGRSGFPFADGVAGPK 240
QY 241 GPAGERSGPGAPKSGPGEAGRPGEAGLFGAKGLTSGPSGPDGKTGTPPAGQDGRP 300
Db 241 GPAGERSGPGAPKSGPGEAGRPGEAGLFGAKGLTSGPSGPDGKTGTPPAGQDGRP 300
QY 301 GPPGPPGARGOAGVMGPPGPKGAAGBPKAGRGVPGPAGVAGKDGAGAGQPPGPA 360
Db 301 GPPGPPGARGOAGVMGPPGPKGAAGBPKAGRGVPGPAGVAGKDGAGAGQPPGPA 360
QY 361 GPAGERGEOGAPSGPFGQLPGPAGPCEACKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
Db 361 GPAGERGEOGAPSGPFGQLPGPAGPCEACKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
QY 421 GPPGERGSGRSGFPFADGVAGPKPAGERSGPGAPKSGPAGRPGEAGLFGAKGLT 480
Db 421 GPPGERGSGRSGFPFADGVAGPKPAGERSGPGAPKSGPAGRPGEAGLFGAKGLT 480
QY 481 GSPGSPGPDGKTGTPPAGQDGRPGPPGARGOAGVMGPPGPKGAAGBPKAGRGVP 540
Db 481 GSPGSPGPDGKTGTPPAGQDGRPGPPGARGOAGVMGPPGPKGAAGBPKAGRGVP 540
QY 541 GPPCAVGPAGKDGAGAGQPPGAPGAGERGEOGAPSGPFGQLPGPAGPPGEAGKPGEQ 600
Db 541 GPPCAVGPAGKDGAGAGQPPGAPGAGERGEOGAPSGPFGQLPGPAGPPGEAGKPGEQ 600
QY 601 GVPGLGAPGSPGAG 616
Db 601 GVPGLGAPGSPGAG 616

RESULT 4
AAY84544
ID AAY84544 standard; protein; 1057 AA.
XX
AC AAY84544;
XX
DT 25-JUL-2000 (first entry)
XX
DE A human collagen 1 (alpha1) protein helical region.
XX
KW Extracellular matrix protein; self aggregation; hydroxylated proline;
KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS Homo sapiens.
XX
PN EP992586-A2.

XX PD 12-APR-2000.
 XX PF 07-OCT-1999; 99EP-00119184.
 XX PR 09-OCT-1998; 98US-00169768.
 XX PA (USSU) US SURGICAL CORP.
 XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX DR WPI; 2000-259138/23.
 XX DR N-PSDB; AAA12503.
 XX PT Production of extracellular matrix proteins containing 4-trans-
 XX PT hydroxyproline results in native self aggregating proteins, useful on
 XX PT medical implants.
 XX PS Example 10; Fig 39A-E; 260pp; English.
 XX CC The specification describes a method for producing an extracellular
 XX CC matrix protein or its fragment. The extracellular matrix protein is
 XX CC capable of self aggregating in a cell which does not ordinarily
 XX CC hydroxylated prolines. The method comprises optimising a nucleic acid
 XX CC sequence for expression in the cell by substitution of codons preferred
 XX CC by that cell for naturally occurring codons not preferred by the cell;
 XX CC incorporating the nucleic acid sequence into the cell; and contacting the
 XX CC cell with a hypertonic growth medium containing at least one amino acid,
 XX CC selected from the group consisting of trans-4-hydroxyproline and 3-
 XX CC hydroxyproline to allow at least one of the amino acids to be assimilated
 XX CC into the cell and incorporated into the extracellular matrix protein. The
 XX CC method may be used to make host cells assimilate and incorporate trans-4-
 XX CC hydroxyproline into proteins. This is especially useful in the
 XX CC recombinant production of proteins such as collagen, fibrinogen and
 XX CC fibronectin whose ability to self aggregate and produce functional
 XX CC proteins depends on the post translational hydroxylation of proline. The
 XX CC method is also useful in studying the structure and function of
 XX CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 XX CC present sequence represents human collagen 1 (alpha1) helical region,
 XX CC which may be produced using the method of the invention
 XX CC Sequence 1057 AA;
 XX CC
 XX CC Query Match 64.1%; Score 2237.5; DB 3; Length 1057;
 XX CC Best Local Similarity 60.2%; Pred. No. 3.2e-130;
 XX CC Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 XX CC
 XX CC 1 GPP-----GEPGFTGLPGPPGRRGSGRFFGADGAGVAGPKGAGERSGPA 48
 XX CC 297 GPPGAGEGKRGAGEGPGTGLPGPPGRRGSGRFFGADGAGVAGPKGAGERSGPA 356
 XX CC 49 GPKGSGEAGRPGEAGLPGAKGLTSGPSGSGDGTGPPGAGQDGRFPDPPGARGQA 108
 XX CC 357 GPKGSGEAGRPGEAGLPGAKGLTSGPSGSGDGTGPPGAGQDGRFPDPPGARGQA 416
 XX CC 109 GVMGPPGPKGAAGEPGKAGRGVPPGAVGAGPKDGEAGAGPPGPPGAGPAGERGQGA 168
 XX CC 417 GVMGPPGPKGAAGEPGKAGRGVPPGAVGAGPKDGEAGAGPPGPPGAGPAGERGQGA 476
 XX CC 169 GSPFGGLPGPAGPGEAGKPGEGQVPGDLGAPGSGPAGE-----PQP----- 212
 XX CC 477 GSPFGGLPGPAGPGEAGKPGEGQVPGDLGAPGSGPAGE-----PQP----- 536
 XX CC 213 -----TGLPGPGERGSGRSGRPPGADGAGVAGPKGAGERSGPA 258
 XX CC 537 GANFAGNDGAKDAGAPGAGSAGLQGMPPGERGAGLPGPKDGRDAGPKGADGSP 596
 XX CC 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGSGDGTGPPGAGQDGRP 300
 XX CC 597 GKDVGRLGTGTPGPPGAGAPGDKGESGSPGAGTARGAPGDRGEPGPPGAGFAGPP 656
 XX CC 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGEPGKAGRGVPPGPA----- 341

Db 657 GADGQAGKGEFGDAGAKGDAGPPGAGPAGPPGPIGNVAGPAGKAGSAGPPGATGPP 716
 QY 342 -----VGPAGKDGDEAGAQGPPGPPAGP-----AGERGQGPAGSGPQGLPGPAGPPGAGKPP 393
 Db 717 GAAGRVGPPGSGNAGPPGPPGAGKGGKGRGTGTPAGRGVGGPPGPPGAGKGGSP 776
 QY 394 GEQGVPGDLGAPGSPGAGEPGTGLPGPPGRRGSGRFFGADGAGVAGPKGAGERSG 453
 Db 777 GADGAPAGPPTGPPQGIAGRGVGLPGQRRGSGRFFGADGAGVAGPKGAGERSG 836
 QY 454 GPAGP---KGSFGEAGRPGEAGLPGAKGLTSGPSGSGDGTGPPGAGQDGRFPGPP 510
 Db 837 GPMGPPGLAGPGESEGREGAPGAEGRGPRDGSPPGAKGDRGTGTPAGPAGPAGPAG 896
 QY 511 GARGGAGVWGRRGPKGAAGEPGKAGRGVPPGPPGAVGAGPKDGE-----AGAQ 558
 Db 897 GPAGKSGDRGTGTGTPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 956
 QY 559 GPPGP-----AGPAGERGEGQGPAGSP---GFQGLPGPAGPAGERGAGKPGEGQVP 603
 Db 957 GPPGPPGSGRQSGSGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1016
 QY 604 GDLGAPGSGPAG 616
 Db 1017 GPPGPPGPPGPPG 1029
 RESULT 5
 ID AAY84403 standard; protein; 1058 AA.
 AC AAY84403;
 DT 12-JUL-2000 (first entry)
 DE Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
 KW Alpha collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
 KW collagen; mussel adhesive protein; bioadhesive.
 OS Homo sapiens.
 FN WO200014201-A1.
 PD 16-MAR-2000.
 PF 07-SEP-1999; 99WO-US020462.
 PR 09-SEP-1998; 98US-0099652P.
 PA (USSU) US SURGICAL CORP.
 PA (PAOL/) PAOLELLA D N.
 PA (GRUS/) GRUSKIN E A.
 PA (BUEC/) BUECHTER D D.
 PI Paolella DN, Gruskin EA, Buechter DD;
 DR WPI; 2000-271051/23.
 DR N-PSDB; AAZ99843.
 PT Incorporating non-natural amino acid into polypeptide, useful e.g. for
 PT production of bioadhesives, by epoxidation or substitution of
 PT dehydroproline residues.
 PS Disclosure; Fig 6; 66pp; English.
 CC The present sequence represents a human type 1 (alpha1) collagen protein.
 CC Peptides derived from the protein were used to demonstrate incorporation
 CC of 3,4-dehydro-L-proline into the peptide, using the method of the
 CC invention. The specification describes a method for the incorporation of
 CC non-natural amino acid into a polypeptide. The method comprises reacting
 CC at least one 3,4-dehydroproline residue in the polypeptide with an
 CC epoxidation reagent from a polypeptide containing at least one 3,4-

CC epoxypyrrolone residue. The method is used for studying the effects of non-natural amino acids on structure and function of polypeptides. The method is also useful for commercial production of collagen or mussel adhesive proteins (which are useful as bioadhesives), and for incorporating a wide variety of groups, including therapeutic ligands and biological probes, into polypeptides

XX
SQ Sequence 1058 AA;

Query Match 64.1%; Score 2237.5; DB 3; Length 1058;
Best Local Similarity 60.2%; Pred. No. 3.2e-130;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPCPPGCGSRGPPGADGVAGPKPAGERGSPGPA 48
DB 298 GPPGPAEGKRGARGEPTGLPPGCGSRGPPGADGVAGPKPAGERGSPGPA 357
QY 49 GPKSGPGEAGRGEAGLPGAKGLTSGSPGPDGTPPGAGDGRPPGPPGARGQA 108
DB 358 GPKSGPGEAGRGEAGLPGAKGLTSGSPGPDGTPPGAGDGRPPGPPGARGQA 417
QY 109 GVMGPPGPKGAAGBFGKAGRGVPGPGAVGAGPKDGEAGAGQPPGAPGAGERGEGQGA 168
DB 418 GVMGPPGPKGAAGBFGKAGRGVPGPGAVGAGPKDGEAGAGQPPGAPGAGERGEGQGA 477
QY 169 GSPGFGQLPGAPGPEAGKGEQGVPGDLAGAPGSPGAGE-----PGP----- 212
DB 478 GSPGFGQLPGAPGPEAGKGEQGVPGDLAGAPGSGARGRPPGGERGVQPPGAPGR 537
QY 213 -----TGLPGPGERGSGSRGPPGADGVAGPKPAGERGSPGPA 258
DB 538 GAGAPGNDGAKGADGAPGSGAGQAGLGMGGERGAAGLPGPKDGRDAGPGADGSP 597
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPDGKTGTPPGAGQDGRP 300
DB 598 GKGVRLGTPTGPPGAPAGDKGESGSPGAGTGAAGPDRGEPGPPGAGFAGPP 657
QY 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGBFGKAGRGVPGPGCA----- 341
DB 658 GADGQPGAKGFPDAGAKGDAGPPGAPGAPGPPGIGNVGAAPGAKGARGSPGATGFP 717
QY 342 -----VGPAGKDXGAGAGPPGAPG-----AGRGEGQAGSPGFGQLPGAPGPEAGKP 393
DB 718 GAAGRVGPPGSGNAGPPGPPGAPGKGGKPRGTGTGAPRGVPPGPPGAPGKGGSP 777
QY 394 GEQGVGDLGAPGSGPAGEPGPTGLPPGGERGSGRPPGADGVAGPKPAGERGSP 453
DB 778 GADGPAGAPGTGPGQIAGRGVGLPQORGERGPPGLPGSPGPKGQSPGASGERGPP 837
QY 454 GPAGP---KGSFGEAGRGEAGLPGAKGLTSGSPGPDGKTGTPPGAGDGRPPGPPGPP 510
DB 838 GPMGPPGLAGPPGSGREGAPGAEPSGPDGSPGAKGDRGTGTPAGPPGAPGAPGAPGV 897
QY 511 GARGQAGVMGPPGPKGAGEPKAGRGVPPGPPGAVGAPGKDXE-----AGAQ 558
DB 898 GPAGKSGDRGTGTGAPGAPGPPGAPGAPGPPGQPRGDKGTGEGIKHGRGFSGLQ 957
QY 559 GPPGP-----AGPAGERGEGQAGSP---GFGQLPGAPGPPGAEKGEQGVPP 603
DB 958 GPPGPPGSGEGPSGAPGPRGPPGSAAGPKDGLNGLPGLPGLPGRGRTGDAGPV 1017
QY 604 GDLGAPGSPGAP 616
DB 1018 GPPGPPGPPGPPG 1030

RESULT 6
AA84540
ID AA84540 standard; protein; 1107 AA.
XX
AC AA84540;
XX
DT 25-JUL-2000 (first entry)

XX Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
DB Extracellular matrix protein; self aggregation; hydroxylated proline;
XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW collagen; fibrinogen; fibronectin; post translational hydroxylation;
KW decorin; chimera.
XX
OS Homo sapiens.
OS Unidentified.
OS Chimeric.
XX
XX Key Location/Qualifiers
FT Misc-difference 858 /note= "Gly encoded by GCT"
XX
XX EP992586-A2.
XX 12-APR-2000.
XX 07-OCT-1999; 99EP-00119184.
XX 09-OCT-1998; 98US-00169768.
XX (USU) US SURGICAL CORP.
XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX WPI; 2000-259138/23.
XX N-PSDB; AAA12500.
XX
XX Production of extracellular matrix proteins containing 4-trans-hydroxyproline results in native self aggregating proteins, useful on medical implants.
XX
XX Claim 24; Fig 18; 260pp; English.
XX
XX The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylated prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a chimeric collagen 1 (alpha1)/decorin protein, which may be produced using the method of the invention

Sequence 1107 AA;

Query Match 64.1%; Score 2237.5; DB 3; Length 1107;
Best Local Similarity 60.2%; Pred. No. 3.3e-130;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPCPPGCGSRGPPGADGVAGPKPAGERGSPGPA 48
DB 297 GPPGPAEGKRGARGEPTGLPPGCGSRGPPGADGVAGPKPAGERGSPGPA 356
QY 49 GPKSGPGEAGRGEAGLPGAKGLTSGSPGPDGKTGTPPGAGDGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGRGEAGLPGAKGLTSGSPGPDGKTGTPPGAGDGRPPGPPGARGQA 416
QY 109 GVMGPPGPKGAAGBFGKAGRGVPPGPPGAVGAPGAKDGEAGAGQPPGAPGAGERGEGQGA 168

Db 717 GAACRVGPPGSGNAGPPGPPAGKGGKGRGTGPPAGRVEVGPVPPGPPGAGKGGSP 776
 Qy 394 GEQVPGDLGAPGSGPAGRGPTGLPDPGPPGGRGPPGADGVAGPKGPPAGRGSP 453
 Db 777 GADGPAGAGPTGPGQIAGQGVVGLPQQRGPPGLPFGSPGKQPPSGASGERGPP 836
 Qy 454 GPAGP---KGSPPGABRGEAGLPGAKLTCSPGSPGDKTGPAGDGRGPPGPP 510
 Db 837 GPMGPPGLAGPPGSGREGAFAGSGSPGDSGPAKGDGTGPPAGPPGAFGAPGV 896
 Qy 511 GARGQAGVMGPPGPKGAAGFPKAGRGVPPGPAVGPAGKDG-----AGAQ 558
 Db 897 GPAGSGDRGTGPPAGPAGVPAGARGPAGPQGPGRGDKGTGQDGRGKGRGFSGLQ 956
 Qy 559 GPPGP-----AGPAGERGEQPPAGSP---GQGLPAGPPGAGKGEQQVP 603
 Db 957 GPPGPPGSGPQSGAGPAGPPGPPGAGKDGGLGPGTGGPPGRTGDAGPV 1016
 Qy 604 GDLGAPGSPGAG 616
 Db 1017 GPPGPPGPPGPPG 1029

RESULT 8
 AAW68485
 ID AAW68485 standard; protein; 1464 AA.
 AC AAW68485;
 XX
 DT 08-DEC-1998 (first entry)
 XX
 DE Human recombinant collagen protein.
 XX
 KW Primer; PCR; amplification; human; collagen; mammal; plant; prosthesis;
 KW cardiac valve; ligament; tendon; skin; gingival implant; perforates;
 KW nerve regeneration; antibiotic; growth factor; cancer; inflammatory;
 KW gelatin; glue; food.
 XX
 OS Synthetic.
 OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Peptide 1..22 "signal peptide"
 FT Protein 23..999
 FT Protein /note= "mature protein"
 FT Cleavage-site 161
 FT Cleavage-site /note= "cleavage site for aminopeptidase"
 FT Cleavage-site 1218
 FT /note= "cleavage site for carboxypeptidase"
 XX W09827202-A1.
 XX
 XX 25-JUN-1998.
 XX PF 17-DEC-1997; 97MO-FR002331.
 XX PR 17-DEC-1996; 96FR-00016224.
 XX (BIOC-) BIOCEM SA.
 XX
 XX Gruber V, Exposito J, Ruggiero F, Conte J, Garrone R, Merot B;
 FI Bournat P;
 XX WPI; 1998-362771/31.
 DR N-PSDB; AAW68485.
 XX
 XX New recombinant nucleic acid for expressing collagen or derivatives in
 PT plants - useful as, e.g. bio-materials and in therapeutic, cosmetic and
 PT odontological compositions.
 XX Disclosure; Fig 7; 138pp; French.
 PS
 XX

CC The invention relates to the production of mammalian collagen in plants.
 CC 2 clones; alpha3 and alpha22, spanning the human collagen type I gene
 CC were isolated from a MG-63 osteosarcoma library. Clone alpha3 contained
 CC 83 bp of the 5' untranslated region and the first 1920 bp of coding
 CC sequence, whereas clone alpha22 contained sequence encoding amino acids
 CC 171-1454 of the protein and around 500 bp of the 3' untranslated region.
 CC The 2 clones were used to generate a number of fragments which were used
 CC to construct composite sequences encoding variant collagen molecules. The
 CC fragments are: (A) containing nucleotides (nt) -4 to 479; (B) containing
 CC TAA upstream of the sequence encoding the PRS (pathogenesis-related
 CC protein S) signal peptide and bases 66-77 from the sequence encoding the
 CC N-terminus of the pro-collagen amino propeptide domain; (C) the whole of
 CC the amino propeptide domain (nt 72-479); (D) all of the amino-telopeptide
 CC domain (nt 474-534) and the N-terminus of the helical region (nt 535-
 CC 1920); (E) the Drall-Bamhi fragment (1709-2808) of alpha22, encoding aa
 CC 567-936 of the central helical domain; (F) the Bamhi-EcoRI (2803-4362)
 CC region of alpha22, encoding aa 936-1192 in the central helical domain
 CC and aa 1193-1454 in the C-propeptide domain; (G) the C-terminus of the C-
 CC propeptide domain (aa 1346-1464) plus stop codons, and (H) as G but
 CC encoding aa 1343-1401 and also including the KDEL motif for retention in
 CC the ER. This sequence represents a recombinant human collagen. The
 CC encoding gene was constructed from fragments (A), (D), (E), (F) and (G).
 CC The recombinant gene is used for expression of mammalian collagen in
 CC plant cells. The transformed plants, their extracts and parts are useful
 CC as biomaterials (haemostatic compresses, sponges or bandages) and in
 CC pharmaceutical, medical, odontological, cosmetic and biotechnological
 CC compositions (e.g. as prostheses for cardiac valves, ligaments or tendons
 CC ; skin substitutes; gingival implants; microcapsules for perfumes; guide
 CC tubes for nerve regeneration; slow release products for antibiotics,
 CC growth factors, anticancer agents or anti-inflammatories; surgical thread
 CC and components of ointments). They are suitable for treating any disorder
 CC related to collagen dysfunction and gelatin, produced from collagen, is
 CC used to produce glues, surgical prostheses and foods
 XX
 XX Sequence 1464 AA;
 Qy Query Match 64.1%; Score 2237.5; DB 2; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 4.1e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 Qy 1 GPP-----GEPGPTGLPFPGERGGSPGPPGADGVAGPKGPPAGRGSPGA 48
 Db 458 GPPGAGEEKRGARGEPGPTGLPFPGERGGSPGPPGADGVAGPKGPPAGRGSPGA 517
 Qy 49 GPKGSPGAEAGRPGEAGLPGAKLTCSPGSPGDKTGPAGDGRGPPGPPGARGQA 108
 Db 518 GPKGSPGAEAGRPGEAGLPGAKLTCSPGSPGDKTGPAGDGRGPPGPPGARGQA 577
 Qy 109 GVMGFPKGAAGEPGKAGRGVPPGPAVGPAGKDGEGAGNQGPPGPPAGRGEGQGA 168
 Db 578 GVMGFPKGAAGEPGKAGRGVPPGPAVGPAGKDGEGAGNQGPPGPPAGRGEGQGA 637
 Qy 169 GSPGFQGLPAGPPGAGKPGEGQVPGDLGAPGSPGAGE-----PGP----- 212
 Db 638 GSPGFQGLPAGPPGAGKPGEGQVPGDLGAPGSPGAGE-----PGP----- 697
 Qy 213 -----TGLPFPGERGGSPGPPGADGVAGPKGPPAGRGSPGPPGARGQA 258
 Db 698 GAGAPGNDGAKGDAGAPGAGPSQAGFLQMPGEGAGLPGPKGDRGDGPKGADSP 757
 Qy 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGPDGKTGPPGPPGARGQA 300
 Db 758 GKDGVRGLTGPFGPPGAGAPGDKGSGSPGPPGAGTGAAGAPGDRGPPGPPGAGGPP 817
 Qy 301 GPPGPPGARGQA-----AGVMGFPKGAAGEPGKAGRGVPPGPPGA----- 341
 Db 818 GADGQPPGAKGEGPDGAKGDAGPPGPPGAGPPGPIGNVGPAGKARGAGPPGATGFP 877
 Qy 342 -----VGPAGKDGEGAGNQGPPGPPGAGP---AGERGEQPPGSPGPPGPPGARGQA 393
 Db 878 GAAGRVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 937
 Qy 394 GEQGVPGDLGAPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 453

Qy	1	GPP-----GEPQDTGLPPGPGRGPGSGRFPQADGVAGPKGPAGERGSGPQA	48
Db	458	GPFGPAGEEGRKARGEPGTTGLPGPPGRGPGSGRFPQADGVAGPKGPAGERGSGPQA	517
Qy	49	GPXGSPGEACRPGCEAGLPGAKGLTGSFSGSPDGKTPPGPAGQDGRPPGPPGARGQA	108
Db	518	GPXGSPGEACRPGCEAGLPGAKGLTGSFSGSPDGKTPPGPAGQDGRPPGPPGARGQA	577
Qy	109	GVNMGFPQKGAAGEPGKAGRGVPGPPGAVGPAGKDGEGAGAAQPPGPGPAGERGEGQPA	168
Db	578	GVNMGFPQKGAAGEPGKAGRGVPGPPGAVGPAGKDGEGAGAAQPPGPGPAGERGEGQPA	637
Qy	169	GSFGFQGLPGPAGPGEAGKPGGQGVPGDILGAPGSPGAGE-----PGP----	212
Db	638	GSFGFQGLPGPAGPGEAGKPGGQGVPGDILGAPGSPGAGERGFPGSGRGVQGPAGPR	697
Qy	213	-----TGLPGPPGRGGPGSGRFPQADGVAGPKGPAGERGSGPAGPKGSP	258
Db	698	GANGAPGNDGAKGDAGAPGSGQAGPLQOMFGERGGAAGLPDTPKDRGDAGPKGADGSP	757
Qy	259	GE-----AGRPGEAGLPGAK-----GLTSGSPSPDGKTPPGPAGQDGRP	300
Db	758	GKDGVRGLTGPIGPPCPGAPAGDKGESGSPGAGPTGARGACPDGRGEPGPPGAPGAPGP	817
Qy	301	GPPEGPGARQ-----AGVMGFPQKGAAGEPGKAGRGVPGPPGA-----	341
Db	818	GADGQPGAKGEPDAGAKGDAGPPGAPGAPGPIGNVGAPKAGARGAGSPGATGFP	877
Qy	342	-----VGPAGKDGEGAGQPPGAPG-----AGERGEGPAGSPGFQGLPGPAGPPGEAGKP	393
Db	878	GAAGRVPVPPGSPNAGPPGPGAGKEGGKPRGETGPARGVEVPPGPPGAGEKSP	937
Qy	394	GEQGVPCDLGAPSPGAPGPGTGLPGPPGRGPGSGRFPQADGVAGPKGPAGERGSP	453
Db	938	GADGPAGACTPPGQGTAGQGVVGLPGQGRGEPFGLPGPSGEPKQKQPSGASGERGPP	997
Qy	454	GPAGP-----KSGPGEAGRPGEAGLPGAKGLTGSFSGSPDGKTPPGPAGQDGRPPGPP	510
Db	998	GPWGPPLGAPPGPESCREGAPGAEPSGRDGSFCAKDRGETPAGPPGAPGAPGAPGV	1057
Qy	511	GARGQAGVMGFPKGNAGEPGKAGRGVPGPPGAVCPAGKDG-----AGAQ	558
Db	1058	GPAGKSDRGETGPAGPVPVGPAGARGPAGPQGRGDKGETGEGQDRGKGHGRGFGSLQ	1117
Qy	559	PPPGP-----AGPAGERGEGPAGSP---CFQGLPGPAGPPGEPAGKPGEGQVP	603
Db	1118	GPFGPSPGEGQPSGASGAPGPPGSPGASAGPKDGLNGLPGPIGPPGPRGRTGDAGPV	1177
Qy	604	GDLGAPGSPGAPG 616	
Db	1178	GPFGPSPGPPG 1190	
RESULT 11			
ADD45051			
ID	ADD45051	standard; protein; 1464 AA.	
AC	ADD45051;		
XX			
XX			
XX	29-JAN-2004	(first entry)	
DE			
XX	Human Protein P02452, SEQ ID NO 10483.		
KW	Human; pain; neuronal tissue; gene therapy;		
KW	spinal segmental nerve injury; chronic constriction injury; CCI;		
KW	spared nerve injury; SNI; Chung.		
OS	Homo sapiens.		
XX			
XX	WO2003016475-A2.		
XX			
XX	27-FEB-2003.		

PF 14-AUG-2002; 2002WO-US025765.
 XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 XX 26-NOV-2001; 2001US-0333347P.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 XX GENBANK; P02452.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 XX preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates a
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1464 AA;
 Query Match 64.1%; Score 2237.5; DB 7; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 4.1e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----GEPDPTGLPDPGRRGSGRPFPGADGVAGPKPAGERGSGPFA 48
 DB 458 GPPFPAGEEKGARGEPGFTGLPDPGRRGSGRPFPGADGVAGPKPAGERGSGPFA 517
 QY 49 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGDPKGTGPPAGQDGRGPPGPGARQQA 108
 DB 518 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGDPKGTGPPAGQDGRGPPGPGARQQA 577
 QY 109 GVMGFPKGAAGEPKKAGRGVPPGAVGAGKDGAGAGQPPGPPGAPAGERGEGQPA 168
 DB 578 GVMGFPKGAAGEPKKAGRGVPPGAVGAGKDGAGAGQPPGPPGAPAGERGEGQPA 637
 QY 169 GSPFGQLGPPGAPGEGAGKPGEQVPGDLAGPSPGAGE-----PGP----- 212
 DB 638 GSPFGQLGPPGAPGEGAGKPGEQVPGDLAGPSPGARGRPFGRGVGPPGAPGR 697
 QY 213 -----TGLPDPGRRGSGRPFPGADGVAGPKPAGERGSGPFA 258
 DB 698 GANGAPGNDGAKGDAGAGFAGSGQAFGLQGMFGERGAGLPGPKDGRGDAGPKGADGSP 757

QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGSPDPKGTGPPGAPGQDGRP 300
 DB 758 GKDGVRGITGPIGPPGAPAGCDKGESGSPGAGTARGAPGDRGEPGPPGAPGAGPP 817
 QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGEPKKAGRGVPPGPGCA---- 341
 DB 818 GADGQFGAKGEPGDAGAGDAGPPGAPGAPGPGPIGNVGFAGKARGSGAPFGATGFP 877
 QY 342 -----VGPAGKDGAGAGGPPGAPG-----AGERGQGAGSPFGQGLGPPGAPGAGK 393
 DB 878 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGETGAPRGPGVGPFGPPGAPAGEKGP 937
 QY 394 GEQGVPGDLGAPGSPGAPGEPGTCGLPDPGRRGSGRPFPGADGVAGPKPAGERGSP 453
 DB 938 GADGAPAGATGTPGQIAGQGVVGLPQGRGERGFPGLPSPGSPGPKGSPGASGERGPP 997
 QY 454 GPAGP---KGSFGEAGRPGEAGLPGAKGLTGSPPGDPKGTGPPGAPGQDGRGPPGPP 510
 DB 998 GPMGPPGLAGPPGSGREGAPGAEFGSPGRDGSFGAKGDRGETGAPGPGAFGAFGAPGV 1057
 QY 511 GARGQAGVMGFPKGAAGEPKKAGRGVPPGPGAVGAPGKDG-----AGAQ 558
 DB 1058 GPAGKSGDRGETGAPGAPGVPVGPAGARGPAGPQGRGDKGTGQDGRGKGRGFGSLQ 1117
 QY 559 GPPGP-----AGPAGERGEGQAGSP---GFGLPGPAGPPGAGKPGEGQGV 603
 DB 1118 GPPGPPGSPGQSGAGSPGAPGPPGSGAGAFKDGGLGLGPIGPPGPRGRTGDAGPV 1177
 QY 604 GDLAGPSPGAPG 616
 DB 1178 GPPGPPGPPGPPG 1190
 RESULT 12
 ADR16800
 ID ADR16800 standard; protein; 1464 AA.
 XX
 AC ADR16800;
 XX
 DT 21-OCT-2004 (first entry)
 XX
 DE Human collagen I alpha1 (I) chain protein.
 XX
 KW T-lymphocyte; AAA; abdominal aortic aneurysm; human; collagen I; therapy.
 XX
 OS Homo sapiens.
 XX
 PN US2004151731-A1.
 XX
 PD 05-AUG-2004.
 XX
 PF 04-FEB-2003; 2003US-00357851.
 XX
 PR 04-FEB-2003; 2003US-00357851.
 XX
 PA (JICH/) JICHA D L.
 XX
 PI Jicha DL;
 XX
 DR WPI; 2004-561474/54.
 XX
 PS Disclosure; SEQ ID NO 1; 24pp; English.
 CC
 CC The present invention relates to a T-lymphocyte derived from abdominal
 CC aortic tissue which are specifically reactive with collagen I, collagen
 CC III or their fragments. The invention is useful to prevent or treat an
 CC abdominal aortic aneurysm (AAA) or rupture in a mammal. The present
 CC sequence is human collagen I alpha1 (I) chain protein. This sequence is
 CC used in the invention.
 XX

SQ Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 8; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 4.1e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPPGCGSGRGGPPGADGVAGPKGAGRGSPGPA 48
 DB 458 GPPGAGEGKRGAGEGPTGLPPGGERGGSGRGGPPGADGVAGPKGAGRGSPGPA 517

QY 49 GPKSGPGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGCARQQA 108
 DB 518 GPKSGPGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGCARQQA 577

QY 109 GVMGPPGKGAAGRPGKAGRGVPPGAVGAGKDGAGAGQDGRGPPGAGRGEGQCPA 168
 DB 578 GVMGPPGKGAAGRPGKAGRGVPPGAVGAGKDGAGAGQDGRGPPGAGRGEGQCPA 637

QY 169 GSPGFGQLPGPAGPGEAGKPGEGVCDLGAAPGSPGAGE-----PQP----- 212
 DB 638 GSPGFGQLPGPAGPGEAGKPGEGVCDLGAAPGSPGAGE-----PQP----- 212

QY 213 -----TGLPPGGERGGSGRGGPPGADGVAGPKGAGRGSPGPA 48
 DB 698 GAGAPGNDGAKGADAGAPGAGSOGAPLQGMPEGGAAGLPGPKGDRGDGPKGADGSP 757

QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPDGKTGPPGAGQDGRP 300
 DB 758 GKGVRGLTGTPIGPPGAGAPGDKGSGSPGAGTGTGARGAPGDRGEPGPPGAGFAGPP 817

QY 301 GPPGPPGARGO-----AGVMGPPGKGAAGRPGKAGRGVPPGPA----- 341
 DB 818 GADGQPGAKGSPGDAGAKGDAGPPGAPGPPGPIGNVAGAPGAKGARGSPGATGFP 877

QY 342 -----VGPAGKDGAGAGQDGRGPPGAGE-----PQP----- 341
 DB 878 GAAGRVGPPGSGNAGPPGPPGKAGKGGKPRGETGPAEGVEGPPGPPGAGEKSP 937

QY 394 GEQGVGPDGLGAPGSPGAGEGPTGLPFPGERGGSGRGGPPGADGVAGPKGAGRGSP 453
 DB 938 GADGAPAGPTGPGQGIAGRGVGLPGQRGREGFPGLPGPSGPPGKQGPSGASGERGPP 997

QY 454 GPAGP---KGSFGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPP 510
 DB 998 GPMGPPGLAGPPGSGREGAPGAEAGSGRDSGPAKGDGRGETGAGPPGAPGAPGVP 1057

QY 511 GARGQAGVMGPPGKGAAGRPGKAGRGVPPGAVGAGKDG-----AGAQ 558
 DB 1058 GPAGKSGDRGTGPAAGPAGVGPAGAPGAPGQDGRGKTGEGDRIKGRGFSGLQ 1117

QY 559 GPPGP-----AGPAGERGEGPAGSP---GFGQLPGPAGPPGAGEAGKPGEGQVPP 603
 DB 1118 GPPGPPGSGGSGSAGSPAGPRGPPGSAAGPKDGLNGLPGPIGPPGPRGRTGDAGPV 1177

QY 604 GDLGAPGSPGAG 616
 DB 1178 GPPGPPGPPGPPG 1190

RESULT 13
 ADRI6425
 ID ADRI6425 standard; protein; 1464 AA.
 XX AC ADRI6425;
 XX DT 21-OCT-2004 (first entry)
 XX DE Human collagen I alpha1 (I) chain protein.
 XX KW T-lymphocyte; AAA; abdominal aortic aneurysm; vaccine; therapy; human;
 XX KW collagen I.
 XX OS Homo sapiens.

XX FN US2004151732-A1.
 XX PD 05-AUG-2004.
 XX PF 04-FEB-2003; 2003US-00358024.
 XX PR 04-FEB-2003; 2003US-00358024.
 XX PA (JICH// JICHA D L.
 XX PA (PELU// PELUSE S.
 XX PI Jicha DL, Peluse S;
 XX DR WPI; 2004-570707/55.
 XX PT Isolated or purified lymphocytes derived from blood in abdominal aortic
 PT aneurysm patients useful for early diagnosis of aneurysms are reactive
 PT with collagen I, collagen III and/or their fragments.
 XX PS Disclosure; SEQ ID NO 1; 22pp; English.
 XX CC The present provides a T-lymphocyte derived from blood in abdominal
 CC aortic aneurysm (AAA) patients which are reactive with collagen I,
 CC collagen III and their fragments. The invention is useful in diagnosis,
 CC prevention and treatment of abdominal aortic aneurysm or rupture in a
 CC mammal. The invention is also useful in the vaccine preparation. The
 CC present sequence is human collagen I alpha1 (I) chain protein.
 XX SQ Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 8; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 4.1e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPPGCGSGRGGPPGADGVAGPKGAGRGSPGPA 48
 DB 458 GPPGAGEGKRGAGEGPTGLPPGGERGGSGRGGPPGADGVAGPKGAGRGSPGPA 517

QY 49 GPKSGPGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGCARQQA 108
 DB 518 GPKSGPGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGCARQQA 577

QY 109 GVMGPPGKGAAGRPGKAGRGVPPGAVGAGKDGAGAGQDGRGPPGAGRGEGQCPA 168
 DB 578 GVMGPPGKGAAGRPGKAGRGVPPGAVGAGKDGAGAGQDGRGPPGAGRGEGQCPA 637

QY 169 GSPGFGQLPGPAGPGEAGKPGEGVCDLGAAPGSPGAGE-----PQP----- 212
 DB 638 GSPGFGQLPGPAGPGEAGKPGEGVCDLGAAPGSPGAGE-----PQP----- 212

QY 213 -----TGLPPGGERGGSGRGGPPGADGVAGPKGAGRGSPGPA 48
 DB 698 GAGAPGNDGAKGADAGAPGAGSOGAPLQGMPEGGAAGLPGPKGDRGDGPKGADGSP 757

QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPDGKTGPPGAGQDGRP 300
 DB 758 GKGVRGLTGTPIGPPGAGAPGDKGSGSPGAGTGTGARGAPGDRGEPGPPGAGFAGPP 817

QY 301 GPPGPPGARGO-----AGVMGPPGKGAAGRPGKAGRGVPPGPA----- 341
 DB 818 GADGQPGAKGSPGDAGAKGDAGPPGAPGPPGPIGNVAGAPGAKGARGSPGATGFP 877

QY 342 -----VGPAGKDGAGAGQDGRGPPGAGE-----PQP----- 341
 DB 878 GAAGRVGPPGSGNAGPPGPPGKAGKGGKPRGETGPAEGVEGPPGPPGAGEKSP 937

QY 394 GEQGVGPDGLGAPGSPGAGEGPTGLPFPGERGGSGRGGPPGADGVAGPKGAGRGSP 453
 DB 938 GADGAPAGPTGPGQGIAGRGVGLPGQRGREGFPGLPGPSGPPGKQGPSGASGERGPP 997

QY 454 GPAGP---KGSFGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPP 510
 DB 998 GPMGPPGLAGPPGSGREGAPGAEAGSGRDSGPAKGDGRGETGAGPPGAPGAPGVP 1057

QY 511 GARGQAGVMGPPGKGAAGRPGKAGRGVPPGAVGAGKDG-----AGAQ 558
 DB 1058 GPAGKSGDRGTGPAAGPAGVGPAGAPGAPGQDGRGKTGEGDRIKGRGFSGLQ 1117

QY 559 GPPGP-----AGPAGERGEGPAGSP---GFGQLPGPAGPPGAGEAGKPGEGQVPP 603
 DB 1118 GPPGPPGSGGSGSAGSPAGPRGPPGSAAGPKDGLNGLPGPIGPPGPRGRTGDAGPV 1177

QY 604 GDLGAPGSPGAG 616
 DB 1178 GPPGPPGPPGPPG 1190

RESULT 13
 ADRI6425
 ID ADRI6425 standard; protein; 1464 AA.
 XX AC ADRI6425;
 XX DT 21-OCT-2004 (first entry)
 XX DE Human collagen I alpha1 (I) chain protein.
 XX KW T-lymphocyte; AAA; abdominal aortic aneurysm; vaccine; therapy; human;
 XX KW collagen I.
 XX OS Homo sapiens.

Db 998 GPMGPPGLAGPESGREGAPGAGSGSPGRDGSFGAKGRGETGTPAGPPGAPGAPGAPGV 1057
 QY 511 GARGQAGVVMGFPKGAAGEGPKAGERGVPPGPGAVGAGKDG-----AGAQ 558
 Db 1058 GPAGKSGDRGETGTPAGPAGPVGPGAGPAGPQGRGDKGETGEQDRIKGRGFSGLQ 1117
 QY 559 GPPGP-----AGPAGERGEQPGASP-----GFGQLPGPAGPGEAGKFGEGQVP 603
 Db 1118 GPPGPPGSGEGGPGSAGPAGPFGPPGSGAGPGRDGLNGLPGTGPGRGTGDAGPV 1177
 QY 604 GDLGAPGSPGAP 616
 Db 1178 GPPGPPGPPGPPG 1190

RESULT 14
 ADR99144
 ID ADR99144 standard; protein; 1464 AA.
 XX
 AC ADR99144;
 DT 02-DEC-2004 (first entry)
 XX
 DE Collagen, type 1, alpha 1, COL1A1, SEQ ID 150.
 XX Cytostatic; breast cancer; cancer; human; Collagen type 1 alpha 1;
 KW COL1A1.
 XX Homo sapiens.
 OS
 XX WO2004078035-A2.
 XX
 PD 16-SEP-2004.
 XX
 PF 27-FEB-2004; 2004WO-US007268.
 XX
 PR 28-FEB-2003; 2003US-0450655P.
 XX
 PA (FARB) BAYER PHARM CORP.
 XX
 PI Eveleigh D, Bigwood D;
 XX
 DR WPI; 2004-653556/63.
 DR N-PSDB; ADR99017.
 XX
 PT Diagnosing breast cancer comprises comparing the level of expression of
 genes or gene products in a first biological sample taken from a patient
 with that in a normal patient sample.
 XX
 PS Claim 3; SEQ ID NO 150; 53pp; English.
 XX

CC The present invention relates to a method (M1) for diagnosing breast
 cancer in a patient. The method comprises comparing the level of
 expression of one or more genes or gene products in a biological sample
 from the patient with that in a normal patient sample, where a difference
 in the gene expression in the first sample compared to that in the second
 sample is a diagnostic of the disease. Also claimed are: method (M2) for
 distinguishing between normal and disease tissues; method (M3) for
 monitoring the response of a breast cancer patient to treatment with an
 anti-cancer agent; method (M4) for identifying a compound for treating
 breast cancer; and an array for distinguishing between normal and disease
 tissues comprising two or more probes corresponding to genes selected
 from ADR98995-ADR99121 or comprising two or more polypeptides selected
 from ADR99122-ADR99248. In M1 and M2 the genes are selected from ADR98995
 -ADR99121 and the gene products are polypeptides selected from ADR98995
 -ADR99248. M1 is useful for diagnosing breast cancer. M2 and the array are
 useful for distinguishing between normal and disease tissue. M3 is useful
 for monitoring the response of a breast cancer patient to treatment with
 an anti-cancer agent. M4 is useful for identifying a compound for
 treating breast cancer. Note: The sequence data for this patent did not
 form part of the printed specification, but was obtained in electronic
 format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 1464 AA;
 Query Match 64.1%; Score 2237.5; DB 8; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 4.1e-130; Indels 117; Gaps 12;
 Matches 441; Conservative 23; Mismatches 152;
 QY 1 GPP-----GEPGPTGLPQPPGGERGPGSGRFGADGVAGPKPAGERGSPGA 48
 Db 458 GPPGAGBEGKRGAGEGFPPTGLPFPGERGPGSRGFGADGVAGPKPAGERGSPGA 517
 QY 49 GPKGSPGEAGRPGEAGLPCAKGLTSGSPGPDGKTGPPGAGODRGPFGPCARQA 108
 Db 518 GPKGSPGEAGRPGEAGLPCAKGLTSGSPGPDGKTGPPGAGODRGPFGPCARQA 577
 QY 109 GVMGPPGPKGAAGEGPKAGERGVFPGAVGPAKGDEAGAQGPPGAPAGERGEGQPA 168
 Db 578 GVMGPPGPKGAAGEGPKAGERGVFPGAVGPAKGDEAGAQGPPGAPAGERGEGQPA 637
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGVPGDILGAPSPSPAGE-----PGP 212
 Db 638 GSPGFQGLPGPAGPGEAGKPGEGVPGDILGAPSPSPAGE-----PGP 697
 QY 213 -----TGLPFPGERGPGSRGFGADGVAGPKPAGERGSPGAPGDRGEPFGPAGP 817
 Db 698 GAGAPGNDGAKGDAGAPGAGPSQAPFLOGMFGERGAAGLPGPKGDRGADGADGSP 757
 QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPDGKTGPPGAGQDGRP 300
 Db 758 GKDGVRGLTGPFGPPGAPAGDGDGSGSPGAPGTGARGAPDRGEPFGPAGPAGP 817
 QY 301 GPPGPPGARGQ-----AGVMGFPQPKGAAGEGPKAGERGVPGPGA-----341
 Db 818 GADGQPKAGEGPDGAKGDAGPPGAPGAPGPPPTGNVGAAPKAGARGAGPPGATGFP 877
 QY 342 -----VCPAKDGEAGAQGPPGAPG-----AGERGEGPAGSPGQGLPGPAGPPGAGKP 393
 Db 878 GAAGRVGPPGPGSNAGPPGPPGAGKEGKGPRGETGTPAGRPGEVGPPGPPGAGEKSP 937
 QY 394 GEQGVPGDILGAPGSPGAPGPTGLPGPGERGPGSRGFGADGVAGPKPAGERGSP 453
 Db 938 GADGAPAGPTFPQGIAGQGVVGLPQGRGERGFPGLPFGSGEPKQKQPSGASGERGPP 997
 QY 454 GPAGP---KGSPCEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPP 510
 Db 998 GPMGPPGLAGPESGREGAPGAGSGSPGRDGSFGAKGRGETGTPAGPPGAPGAPGAPGV 1057
 QY 511 GARGQAGVVMGFPKGAAGEGPKAGERGVPPGPGAVGAGKDG-----AGAQ 558
 Db 1058 GPAGKSGDRGETGTPAGPAGPVGPGAGPAGPGRDGLNGLPGTGPGRGTGDAGPV 1117
 QY 559 GPPGP-----AGPAGERGEQPGASP-----GFGQLPGPAGPGEAGKFGEGQVP 603
 Db 1118 GPPGPPGSGEGGPGSAGPAGPFGPPGSGAGPGRDGLNGLPGTGPGRGTGDAGPV 1177
 QY 604 GDLGAPGSPGAP 616
 Db 1178 GPPGPPGPPGPPG 1190

RESULT 15

AAE02535

ID AAE02535 standard; protein; 1449 AA.

XX

AC AAE02535;

XX

DT 10-AUG-2001 (first entry)

XX

DE Porcine alpha1(I) collagen.

XX

KW Porcine; alpha1(I) collagen; gelatin; cytostatic; viral infection;
 KW Pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
 KW Medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;
 KW Rheumatoid arthritis; beverage; photographic application.

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description	
1	2237.5	64.1	1464	4	US-09-331-347C-21	Sequence 21, Appl	
2	2238.5	63.9	1341	3	US-08-963-825-18	Sequence 18, Appl	
3	2238.5	63.9	1341	3	US-09-500-811-18	Sequence 18, Appl	
4	2238.5	63.9	1341	3	US-09-570-573-18	Sequence 18, Appl	
5	2238.5	63.9	1341	3	US-09-548-608-18	Sequence 18, Appl	
6	2224.5	63.8	1461	4	US-09-585-887-9	Sequence 9, Appl	
7	2224.5	63.8	1461	4	US-09-289-578-9	Sequence 9, Appl	
8	2208.5	63.3	1057	3	US-08-931-820-1	Sequence 1, Appl	
9	2191	62.8	822	3	US-09-219-849-49	Sequence 49, Appl	
10	1955	56.0	1017	4	US-08-468-996-10	Sequence 10, Appl	
11	1943	55.7	1060	3	US-08-931-820-3	Sequence 3, Appl	
12	1943	55.7	1418	3	US-08-963-823-20	Sequence 20, Appl	
13	1943	55.7	1418	3	US-09-010-999-1	Sequence 1, Appl	
14	1943	55.7	1418	3	US-09-500-811-20	Sequence 20, Appl	
15	1943	55.7	1418	3	US-09-570-573-20	Sequence 20, Appl	
16	1943	55.7	1418	3	US-09-548-608-20	Sequence 20, Appl	
17	1936	55.5	1442	2	US-08-316-650-12	Sequence 12, Appl	
18	1936	55.5	1442	5	PCT-US95-02251-12	Sequence 12, Appl	
19	182.5	54.0	595	3	US-09-219-849-48	Sequence 48, Appl	
20	182.5	54.0	595	3	US-09-219-849-50	Sequence 50, Appl	
21	182.5	54.0	1064	1	US-08-642-255-62	Sequence 62, Appl	
22	1853	53.1	1057	3	US-08-931-820-4	Sequence 4, Appl	
23	1837	52.7	720	3	US-09-219-849-4	Sequence 4, Appl	
24	1837	52.7	777	1	US-08-642-255-53	Sequence 53, Appl	
25	1833.5	52.6	1078	3	US-08-963-823-21	Sequence 21, Appl	
26	1833.5	52.6	1078	3	US-09-500-811-21	Sequence 21, Appl	
27	1833.5	52.6	1078	3	US-09-570-573-21	Sequence 21, Appl	

Query Match	64.1%	Score	2237.5	DB	4	Length	1464
Best Local Similarity	60.2%	Pred. No.	1.3e-133				
Matches	441	Conservative	23	Mismatches	152	Indels	117
Gaps	12						
QY	1	GPP-----	-----GEPGPTGLPGPPGRRGGPSRGFP	GADGVAGPKGPAGRRGSPGA	48		
DB	458	GPFGPAGEBKRGARGEPGPTGLPGPPGRRGGPSRGFP	GADGVAGPKGPAGRRGSPGA	517			
QY	49	GPXGSPCEAGRPRGEAGLPGAKGLTGTSPGSPGDKTGP	PGPAGQDGRPGPPGPPGARGQA	108			
DB	518	GPXGSPCEAGRPRGEAGLPGAKGLTGTSPGSPGDKTGP	PGPAGQDGRPGPPGPPGARGQA	577			
QY	109	GVNMGFPFGPKAAAGEPKKAGSRGVDPGPBPAVGP	PAGKDGEGAAQGPDPGAPGABRGEGQPA	168			
DB	578	GVNMGFPFGPKAAAGEPKKAGSRGVDPGPBPAVGP	PAGKDGEGAAQGPDPGAPGABRGEGQPA	637			
QY	169	GSFGQFLGCPAGPPCEAGKPGGQGVPGDLGAPDQSP	GAGE-----PGP----	212			
DB	638	GSFGQFLGCPAGPPCEAGKPGGQGVPGDLGAPDQSP	GAGE-----PGP----	697			
QY	213	-----TGLPGPPBRRGCGPSRGFP	GADGVAGPKGPAGRRGSPGAPKGGSP	258			
DB	698	GANGAPGNDKAGDKDAGAPNPGSQAPGLQCMF	GERGAAGLPQPKDGRGDAGPKGADGSP	757			
QY	259	GE-----AGRPGEAGLPGAK-----	GLTGPSGSPGDKGTGPPGAGQDGRP	300			
DB	758	GKDGVRLGTGPIGPPPGPAGAPGDKGESSGSPAGPT	GARGAPGDRCEPGPPGPPGAPGAGPP	817			
QY	301	GPFGPPGARGQ-----	AGVMGFFGPKGAAGEPKKAGSRGVDPGP	341			

Db	818	GADQOPGAKGEPGDAGKAGDAGPCGAPGPGPIGNVGAPGAKGARGSAGPPCATGFP	877
Qy	342	-----VGPAGKDXGAGAOQPPGPAGP---AGERGEQGPAGSPFGQGLPGPAGPPGEAGKP	393
Db	878	GAAGRVGPFGPSGNAGPGPGPGAGKEGKGPRGETGPAGRPFGEVPPPPGPAGPAGKGP	937
Qy	394	GEQGVPGDLGAPGSPGAGEPPTGLPGPPGERCGPSRGPFGADGVAGPKGPAGERGP	453
Db	938	GADGPAGAPGTPPGQIAGQGVVGLPGQRGERFPGLPGPSGEPKQGPSGASGERGPP	997
Qy	454	GPAGP---KSPGSEAGRPGEACLGAKGLTSGSPGPDGKTGTPPGPAGQDGRGPPGP	510
Db	998	GPMGPPGLAGPPGSGREGAPGAEPSFGSDGSPGAKGDRGETGAPGPPGAPGAPGAPGV	1057
Qy	511	GARQAGVMGPPGPKGAAGEPKAGERVCPGPGAVGPAKDXE-----AGAQ	558
Db	1058	GPAGKSGDRBETGPAGPAGVPVGPAGAKGPAGPQGPGRDKGTGQGDRIKGHRFGSLQ	1117
Qy	559	GPFGP-----AGPAGERGEQGPAGSP---GFQGLPGPAGPPGEAGKPGEGQVP	603
Db	1118	GPFGPPGSPCEQGPSGASGPAGPRGPGPSAGAPKDKLGLNGLPGFIPGPPGRGTGDAGPV	1177
Qy	604	GDLGAPGSPGPAG	616
Db	1178	GPFGPPGPPGPPG	1190

RESULT 2

```

US-08-363-825-18
; Sequence 18, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of

```

```
?  
? ZIP= 10022  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patent In Release #1.0, Version #1.25  
? CURRENT APPLICATION DATA:  
? APPLICATION NUMBER: US/08/963,925  
?
```

/ PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US/08/187,319
 / FILING DATE: 21-JAN-1994
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Gogoris, Adda C
 / REGISTRATION NUMBER: 29,714
 / REFERENCE/DOCKET NUMBER: 4305/087001
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 213-527-7700
 / TELEFAX: 212-753-6237

```

/ ILEUK: 236687
/ INFORMATION FOR SEQ ID NO: 18:
/
/ SEQUENCE CHARACTERISTICS:
/
/ LENGTH: 1341 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ ORIGINAL SOURCE:
/

```

```

; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-08-963-825-18

```

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
Best Local Similarity 60.0%; Pred. NO. 4.6e-133;
Matches 440; Conservative 23; Mismatches 153; Indels 117;

QY	1	GPP-----	GBPGPTGTLPGPPGRRGGPSGRGFP	GADGVAGPKGAPGERSGPGA	48
Db	334	GPFPAGBEGKRGARGBP	GTGLTGPFPGERGGPSRGFP	GCADGVAGPKGAPGERSGPGA	393
QY	49	GPKGSPGAGRPAGLPGAK	GLTGPSGSPDGKTPGPCPAGODRRP	PGPPPGARGQA	108
Db	394	GPKGSPGAGRPAGLPGAK	GLTGPSGSPDGKTPGPCPAGODRRP	PGPPPGARGQA	453
QY	109	GVWGPPGPKGAAGFPKGAG	RGVPGPVGAPGAKDGEAGAGOP	PGPAPGACGERGEOGPA	168
Db	454	GVWGPPGPKGAAGFPKGAG	RGVPGPVGAPGAKDGEAGAGOP	PGPAPGACGERGEOGPA	513
QY	169	SGPFGQLPGPAGPBGAG	KPGBQGVPGDLGAPGPSGPAGE	-----PGP----	212
Db	514	SGPFGQLPGPAGPBGAG	KPGBQGVPGDLGAPGPSGPAGE	-----PGP----	573
QY	213	-----	TGLPGPPGRRGGPSGRGFP	GCADGVAGPKGAPGERSGPGA	258
Db	574	GANGAPNDGAKGDAGAP	AGPAGSQAGPLQGMFGERGAAGL	PGKDRDAGPKGADGSP	633
QY	259	GE-----	AGRPGBAGLPGAK-----	GLTGPSGSPDGKTPGPCPAGODRRP	300
Db	634	KDGVRGLTGPIPGPPGAG	PGDKBSGSPGAPGTGARGA	PDGRGEPGPPGAPGAGPP	693
QY	301	GPFPGPGARGQ-----	AGVMFPGPKGAAGFPGKAG	RGVPGPPGA-----	341
Db	694	GADQCPGAKGEPGDAGAK	GADAGPPGPAGPAGPPGPI	IGNVGAFCAGKARGSAGPPGATGFP	753
QY	342	-----	VGPAGKDGEAGAGPPGPAGP	---AGRGEOGPAGSPGQGLPGPAGPPG	393
Db	754	GAGRVPGPSPGNAGPPG	PPGAPGKXGKGKPRGETGP	APRGFEVGPFGPPGPAKEGKSP	813
QY	394	GEQGVPGDLAGPSPGAPG	BPPTGLPGPPGRRGGPSRGFP	GDGVAGPKGAPGACGERSP	453
Db	814	GADGPAGAPGTPPGTIG	AGRGVVGLPGQGRGRRGFP	GLPGSPGEPKQGPSGASGERPP	873
QY	454	GPAGP---KSGSPGAGRP	BAGLPGAKGLTGPSGSPDGKTP	PGPAGQDGRGPPGPP	510
Db	874	GPMPGPGLAGPPGESR	EGACAPGAEGSPGRDGSF	GKADRGETGPAGPPGAGXGAPGV	933
QY	511	GARGQAGVMGFPKGAAG	EPKAGRGVPGPPGAVGPA	KDGE-----AGAQ	558
Db	934	GPAGKSGDRGETGPAG	PAGPVGPGARGPAGPQG	PRGDKGETGEOGDRGIXKHGRFSG	993
QY	559	GPFGP-----	AGPAGRRGOGPAGSP---	GFOGLPGPAGPCEAGKGEQVPP	603
Db	994	GPFPGPSPEQGPSGAS	GAPGPRPPGSGAPGKDGL	NGLPGPIGPPGPRGRTGDAGPV	1053
QY	604	GDLGAPGPSGAP	616		
Db	1054	GPFGPPGPPGPG	1066		

RESULT, T 3

US-09-500-811-18
Sequence 18, Application US/09500811
Patent No. 632314
GENERAL INFORMATION:
APPLICANT: Ovast, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for
in Body Fluid
TITLE OF INVENTION: Method and U
TITLE OF INVENTION: Disorders A

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/500,811
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-500-811-18

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
Best Local Similarity 60.0%; Pred. No. 4,6e-133;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;
QY 1 GPP-----GEPGTLGPPGGERGSGRPFPGADGVAGPKGAGERSGPPA 48
DB 334 GPPGAGEGKRGARGEPGTLGPPGGERGSGRPFPGADGVAGPKGAGERSGPPA 393
QY 49 GPKSGFAGRPGEGAGLFGAKGLTSGSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 108
DB 394 GPKSGFAGRPGEGAGLFGAKGLTSGSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 453
QY 109 GVMGPPGKGAAGEPKKAGRGVPCGAVGAGDGEAGAGPPGPPGAGERGEOGPA 168
DB 454 GVMGPPGKGAAGEPKKAGRGVPCGAVGAGDGEAGAGPPGPPGAGERGEOGPA 513
QY 169 GSPGQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSGPAGE-----PGP---- 212
DB 514 GSPGQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSGGARGRGPPGGERGVQPPGAPGR 573
QY 213 -----TGLPGPPGERGSGRPFPGADGVAGPKGAGERGSGPPGPKGSP 258
DB 574 GANPAGNDKAGDAGAPGAGSGAGLQGMFGERGAAGLFGPKGDRGDAGPKGADGSP 633
QY 259 GE-----AGRGEAGLPGAK-----GLTSGSGSPGPDGKTGPPGAGQDGRP 300
DB 634 GKDGVRLGTGTPPGPPGAPGDKGESGSPGAGTGAAGPDRGEPGPPGPPGAPGPP 693
QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGEPKKAGRGVPCGAGP----- 341
DB 694 GADGQPKAGFEGDAGAKGDAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 753
QY 342 -----VGPAGKDGEGACAGQPPGAPG-----AGRGSGQPGAGSPGFGQLPGPAGPGEAGKP 393

DB 754 GAAGRVGPPGPPSGNAGPPGPPGPKGEGKGRGETGPPAGRPGEVPPGPPGAGEXGSP 813
QY 394 GEOGVPGDLGAPSGSPAGBPPGTLGPPGGERGSGRPFPGADGVAGPKGAGERGSP 453
DB 814 GADGPAGAPGTGPPGQIAGQGVVGLPQGRGERGFPGLPSPGSPGKQSPGSGASGERGPP 873
QY 454 GPAGP-----KSGPGEAGRPGEAGLFGAKGLTSGSGSPGPDGKTGPPGAGQDGRPGPP 510
DB 874 GPMGPPGLAGPPGESGREGAPGAEPSGRDGSPPKAGKDRGETGPPAGPPGAXGAGPVP 933
QY 511 GARGQAGVMGPPGPKGAAGEPKKAGRGVPCGAVGAGDGE-----AGAQ 558
DB 934 GPACKSGDRGETGPPAGPAGVPGAGARGPAGPQGRGDKGTGQDGRGKIGKRGHGFSLQ 993
QY 559 GPPGP-----AGPAGERGEOGPPGSP--GFGGLPGPAGPPGEGAGKPGEGQVPP 603
DB 994 GPPGPPGSGPGEQSGASGAPGRGPPGSGAGAPGKGLGGLPGIPGPPGPRGRTGDAGVP 1053
QY 604 GDLGARGSPGAP 616
DB 1054 GPPGPPGPPGPPG 1066

RESULT 4

US-09-570-573-18
Sequence 18, Application US/09570573
Patent No. 6342361
GENERAL INFORMATION:

APPLICANT: Oviast, Per
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/570,573
APPLICATION NUMBER: US/09/570,573
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)

US-09-570-573-18

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
Best Local Similarity 60.0%; Pred. No. 4.6e-133;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

```
QY 1 GPP-----GEGPTGLPPPPGRRGGPSRGGADGAGVAGKPGAGERSGPGPA 48
Db |||
334 GPPGAGGKRGARGEPTGLPPGRRGGPSRGGADGAGVAGKPGAGERSGPGPA 393
QY 49 GPKSGFAGRPGAEAGLCAKGLTSGPSGPDGKTGTPGAGDGRPPGPPGARGQA 108
Db |||
394 GPKSGFAGRPGAEAGLCAKGLTSGPSGPDGKTGTPGAGDGRPPGPPGARGQA 453
QY 109 GVMGFPKGAAGEPKKAGRGVPPGAVGAPGAGKDEAGAGQGGPPGAGRGEGGPA 168
Db |||
454 GVMGFPKGAAGEPKKAGRGVPPGAVGAPGAGKDEAGAGQGGPPGAGRGEGGPA 513
QY 169 GSPGFQGLPGAGPGEAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
Db |||
514 GSPGFQGLPGAGPGEAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
QY 213 -----TGLPPGRRGGPSRGGADGAGVAGKPGAGERSGPGAGPR 573
Db |||
574 GANGAPNDGAKGDAGAPGAPGSGQAGPLGQMPGEGGAAGLPGFKDGRDAGPKGADGSP 633
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGPDGKTGTPGAGDGRPP 300
Db |||
634 GKDGVRGLTGPIPPGPPGAPAGDKGESGSPGAPGTGARGAPGDRCEPGPPGAGFAGP 693
QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGEPKKAGRGVPPGPA----- 341
Db |||
694 GADGQPKAGEPGDCAKGDAGPPGAGPAGPPGPIGNVGAAPKAGARSAGPPGATGFP 753
QY 342 -----VGPAGKDEAGAGQGGPPGAGP-----AGRGEGQAGSPGQGLPGAP 393
Db |||
754 GAAGRVPKPSGNAGPPGPPGAGKGGKGRGTGTPAGRGVEGVPFPPGAGEKSP 813
```

RESULT 5

US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Oviatt, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC

```
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/548,608  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1341 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (I)  
US-09-548-608-18
```

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
Best Local Similarity 60.0%; Pred. No. 4.6e-133;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

```
QY 1 GPP-----GEGPTGLPPPPGRRGGPSRGGADGAGVAGKPGAGERSGPGPA 48
Db |||
334 GPPGAGGKRGARGEPTGLPPGRRGGPSRGGADGAGVAGKPGAGERSGPGPA 393
QY 49 GPKSGFAGRPGAEAGLCAKGLTSGPSGPDGKTGTPGAGDGRPPGPPGARGQA 108
Db |||
394 GPKSGFAGRPGAEAGLCAKGLTSGPSGPDGKTGTPGAGDGRPPGPPGARGQA 453
QY 109 GVMGFPKGAAGEPKKAGRGVPPGAVGAPGAGKDEAGAGQGGPPGAGRGEGGPA 168
Db |||
454 GVMGFPKGAAGEPKKAGRGVPPGAVGAPGAGKDEAGAGQGGPPGAGRGEGGPA 513
QY 169 GSPGFQGLPGAGPGEAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
Db |||
514 GSPGFQGLPGAGPGEAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
QY 213 -----TGLPPGRRGGPSRGGADGAGVAGKPGAGERSGPGAGPR 573
Db |||
574 GANGAPNDGAKGDAGAPGAPGSGQAGPLGQMPGEGGAAGLPGFKDGRDAGPKGADGSP 633
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGPDGKTGTPGAGDGRPP 300
Db |||
634 GKDGVRGLTGPIPPGPPGAPAGDKGESGSPGAPGTGARGAPGDRCEPGPPGAGFAGP 693
QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGEPKKAGRGVPPGPA----- 341
Db |||
694 GADGQPKAGEPGDCAKGDAGPPGAGPAGPPGPIGNVGAAPKAGARSAGPPGATGFP 753
QY 342 -----VGPAGKDEAGAGQGGPPGAGP-----AGRGEGQAGSPGQGLPGAP 393
Db |||
754 GAAGRVPKPSGNAGPPGPPGAGKGGKGRGTGTPAGRGVEGVPFPPGAGEKSP 813
```



```
QY 394 GEGVPCDLGAPGSPGAGEGPTGLPCPPGCGSGRSGPPGADGVAGPKGPAGERGSP 453
Db 814 GADGAPAGAPGTPGPGIAGQGVVGLPQQRGERGFPGLPGSGEPGKQSPGASGERGPP 873
QY 454 GPAGP---KGSFGEAGRGEAGLPCAKLGTGSPGSPGDKTGPDPGAGQDGRGPPGPP 510
Db 874 GPMGPPGLAGPPGSGREGAPGAEGRGSPGDCSPGAKGDRGETGPPAGPPGAXGAPGVP 933
QY 511 GARGQAGVMGPPGPKGAAGEFGKAGRGVPPGPPGAGVAGPKGDGE-----AGAQ 558
Db 934 GPAGKSGDRGETGTPAGPAGPVGPGAGPAGPQGRGDKGTGEGQDGRGKRGHGFSLQ 993
QY 559 GPPGP-----AGPAGERGEGQAGSP---GFQGLPGPAGPPGAEAGKPGEGQVP 603
Db 994 GPPGPPGSGQSGASGAPGAPGPPGSGAPGKDGKGLNGLPPIGPPPRGRTGDAGPV 1053
QY 604 GDLGAPGSPGAG 616
Db 1054 GPPGPPGPPGPPG 1066

RESULT 6
US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;
Best Local Similarity 59.9%; Pred. No. 8.8e-133;
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPCPPGCGSGRSGPPGADGVAGPKGPAGERGSPGPA 48
Db 455 GPPGAGEGKRGAGEGPTGLPCPPGCGSGRSGPPGADGVAGPKGPAGERGSPGPA 514
QY 49 GPKGSPGAGRGPBAGLPCAKLGTGSPGSPGDKTGPDPGAGQDGRGPPGPPGARGQA 108
Db 515 GPKGSPGAGRGPBAGLPCAKLGTGSPGSPGDKTGPDPGAGQDGRGPPGPPGARGQA 574
QY 109 GVMGFPKGAAGEFGKAGRGVPPGPPGAGVAGPKGDGEAGAGQPPGPPGAGERGEGQA 168
Db 575 GVMGFPKGAAGEFGKAGRGVPPGPPGAGVAGPKGDGEAGAGQPPGPPGAGERGEGQA 634
QY 169 GSPGFQGLPGPAGPPGAEAGKPGEGQVPCDLGAPGSPGAGE-----GPP----- 212
Db 635 GSPGFQGLPGPAGPPGAEAGKPGEGQVPCDLGAPGSPGAGERGPPGPPGPPGPPGPP 694
QY 213 -----TGLPCPPGCGSGRSGPPGADGVAGPKGPAGERGSPGPPGPPGPPGPP 258
Db 695 GAGPAGNDGAKGDAGAPGAGQAPGLQMPGGERGAGLPGPKGDRGDGAPKGDGSP 754
```

```
QY 259 GE-----AGRPCEAGLPGAK-----CLTGSFGSPGSPGDKTGPDPGAGQDGRP 300
Db 755 GKQGVRLGTPIPPPGPAGAPGDKGSGSPGSPGAGPTGARGAPGDRGEPGPPGPPGPPG 814
QY 301 GPPGPPGARGQ-----AGVMGFFGPKGAAGEFGKAGRGVPPGPPGPA----- 341
Db 815 GADQPCAKGEPGDAGAKGDAGPPGPPGAGPAGPPGPIGNVGAPGAKGARGAGSAGPPGATGFP 874
QY 342 -----VCPAGKDEAGAGQPPGAGP---AGERGEGQAGSPGPGQGLPGPAGPPGAEAGK 393
Db 875 GAAGRVPGPPGSGNAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 934
QY 394 GEGVPCDLGAPGSPGAGEGPTGLPCPPGCGSGRSGPPGADGVAGPKGPAGERGSP 453
Db 935 GADGAPAGAPGTPGPGIAGQGVVGLPQQRGERGFPGLPGSGEPGKQSPGASGERGPP 994
QY 454 GPAGP---KGSFGEAGRGPBAGLPCAKLGTGSPGSPGDKTGPDPGAGQDGRGPPGPP 510
Db 995 GPMGPPGLAGPPGSGREGAPGAEGRGSPGDCSPGAKGDRGETGPPAGPPGAXGAPGVP 1054
QY 511 GARGQAGVMGPPGPKGAAGEFGKAGRGVPPGPPGAGVAGPKGDGE-----AGAQ 558
Db 1055 GPAGKSGDRGETGTPAGPAGPVGPGAGPAGPQGRGDKGTGEGQDGRGKRGHGFSLQ 1114
QY 559 GPPGP-----AGPAGERGEGQAGSP---GFQGLPGPAGPPGAEAGKPGEGQVP 603
Db 1115 GPPGPPGSGQSGASGAPGAPGPPGSGAPGKDGKGLNGLPPIGPPPRGRTGDAGPV 1174
QY 604 GDLGAPGSPGAG 616
Db 1175 GPPGPPGPPGPPG 1187

RESULT 7
US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;
Best Local Similarity 59.9%; Pred. No. 8.8e-133;
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPCPPGCGSGRSGPPGADGVAGPKGPAGERGSPGPA 48
Db 455 GPPGAGEGKRGAGEGPTGLPCPPGCGSGRSGPPGADGVAGPKGPAGERGSPGPA 514
QY 49 GPKGSPGAGRGPBAGLPCAKLGTGSPGSPGDKTGPDPGAGQDGRGPPGPPGARGQA 108
Db 515 GPKGSPGAGRGPBAGLPCAKLGTGSPGSPGDKTGPDPGAGQDGRGPPGPPGARGQA 574
QY 109 GVMGFPKGAAGEFGKAGRGVPPGPPGAGVAGPKGDGEAGAGQPPGPPGAGERGEGQA 168
Db 575 GVMGFPKGAAGEFGKAGRGVPPGPPGAGVAGPKGDGEAGAGQPPGPPGAGERGEGQA 574
QY 109 GVMGFPKGAAGEFGKAGRGVPPGPPGAGVAGPKGDGEAGAGQPPGPPGAGERGEGQA 168
```


Db	575	GWVGFPGPKGAAGEPKKAGERGVP	GGPAGV	PAGKDG	EAAG	AGQPP	CGPAG	CGPAG	GERGE	QGP	634		
Qy	169	GSFPGFQGLPCGPAGP	CGEACK	PGEQ	VGPD	GLGAP	SGPAGE	-----	PG	-----	212		
Db	635	GSFPGFQGLPCGPAGP	CGEACK	PGEQ	VGPD	GLGAP	SGPAGE	-----	PG	-----	694		
Qy	213	-----	TGLPGPP	GERG	PGSR	PGF	GADGV	AGPK	PAGER	SGSP	CGPAG	258	
Db	695	GANGAP	GNDAK	GDAG	APAG	PSQ	AGPLQ	GNP	GERGA	GLP	PKDK	GDAG	754
Qy	259	GE	-----	AGR	PGEAGL	PGAK	-----	GLT	SPGS	PGDK	TGPP	CGPAG	300
Db	755	GKDVG	RLT	GT	PTG	PGP	AGAP	GD	KGES	PG	AGT	GARG	814
Qy	301	GP	PPG	PARG	Q	-----	AGV	MP	PGPK	GA	ER	PKA	341
Db	815	GADQ	PGAK	GP	GDAG	KDAG	PPCP	AG	PPG	PI	GNV	APGAK	874
Qy	342	-----	VGPAG	KDG	EAAG	AGQPP	GPAGP	-----	AGER	GEQ	PAGS	PGFQ	934
Db	875	GAAC	R	VGP	PGPS	GNAG	PGPP	GP	AGK	EGCK	PR	GTG	994
Qy	394	GEQ	V	PD	GLGAP	SGPAGE	PGT	CL	PGP	GER	CGPS	RG	453
Db	935	GAD	P	AG	AGT	PTG	QIQ	AGQ	V	GV	LPG	QR	994
Qy	454	GPAG	---	KGS	P	GB	AGR	PGEAGL	PCAK	GLT	SGPS	PGD	510
Db	995	GP	M	PP	GLAG	PPG	BS	REG	APGA	EGS	PCRD	CS	1054
Qy	511	GARQ	AG	V	MP	PGPK	GAAGE	PKA	GER	GV	PPGP	PGAV	558
Db	1055	GP	AKS	GD	RET	GT	PAG	PAGV	PGV	PG	ARG	PAGP	1114
Qy	559	GP	PP	-----	AG	PAGE	GEQ	PAGS	---	CFQ	GL	CPAG	603
Db	1115	GP	PG	PS	PGEQ	PS	AGS	AGP	AGP	GR	PG	PG	1174
Qy	604	GD	LAG	PP	SG	PG	AG	616	-----	AG	PA	GER	1187

RESULT 8

RESOL 8
US-08-931-820-1
/ Sequence 1, Application US/08931820
/ Patent No. 6010863
/ GENERAL INFORMATION:
/ APPLICANT:
/ TITLE OF INVENTION: Assay for collagen degradation
/ NUMBER OF SEQUENCES: 4
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/931,820
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: EP 96202596.1

RESULT 9

RECORD. 3
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUNSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W. T.
; APPLICANT: WIND, RICHELE D.

APPLICANT: VAN DEN BOSCH, TANJA J.
TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
TITLE OF INVENTION: PREPARATION THEREOF
FILE REFERENCE: 2728-2
CURRENT APPLICATION NUMBER: US/09/219,849
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patent in ver. 2.1
SEQ ID NO 49
LENGTH: 822
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match 62.8%; Score 2191; DB 3; Length 822;
Best Local Similarity 56.9%; Pred. No. 7e-131;
Matches 436; Conservative 22; Mismatches 150; Indels 150; Gaps 13;

Qy 1 GPPGPPGTLPGPPGE-----RGPGSRGPPGADGVAGPKGPPAGRGSPGPA 48
Db 12 GPPGAPGQGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 71
Qy 49 GPKGSPGAGRR-----GEAGLPKAGLGTGSPGPPGPPGPPGPPGPPGPPG 93
Db 72 GARGLPGLTAGLPGMKGHRGSLDCAKGDAGPAGPKGPPGPPGPPGPPGPPG 131
Qy 94 GRPPGPPGARGQGVGPPGPKAAGE-----PGKAGRGVPPGPPGPPGPPG 147
Db 132 GRPPGPPGARGQGVGPPGPKAAGE-----PGKAGRGVPPGPPGPPGPPG 191
Qy 148 GAQGPSPGAGRGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 192
Db 192 GPPGPPGPPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 251
Qy 193 GVPGLDGLAP-----GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 225
Db 252 GNSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 311
Qy 226 GSRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 285
Db 312 GSRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 371
Qy 286 GRTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 345
Db 372 GRTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 431
Qy 346 GKDGEAGQGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 405
Db 432 GKDGEAGQGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 491
Qy 406 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 435
Db 492 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 551
Qy 436 GADGVAGPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 486
Db 552 GERGAAGLPKGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 611
Qy 487 GPD-----GKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 522
Db 612 GFTGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 671
Qy 523 GPKGAAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 570
Db 672 GPIGNVAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 731
Qy 571 GQGGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 616
Db 732 GETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 777

RESULT 10
US-08-468-996-10
Sequence 10, Application US/08468996
Patent No. 6645504
GENERAL INFORMATION:
APPLICANT: Weiner, Howard
APPLICANT: Miller, Ariel
APPLICANT: Zheng, Zheng
APPLICANT: Ahmad, Al-Sabbagh
TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION OF
TITLE OF INVENTION: GLUCAGON
FILE REFERENCE: 1010/16959-US3
CURRENT APPLICATION NUMBER: US/08/468,996
CURRENT FILING DATE: 2003-02-07
PRIOR APPLICATION NUMBER: US 07/843,752
PRIOR FILING DATE: 1992-02-28
PRIOR APPLICATION NUMBER: US 07/460,852
PRIOR FILING DATE: 1990-02-21
PRIOR APPLICATION NUMBER: US 07/596,936
PRIOR FILING DATE: 1990-10-15
PRIOR APPLICATION NUMBER: US 07/065,734
PRIOR FILING DATE: 1987-06-24
PRIOR APPLICATION NUMBER: US 07/454,486
PRIOR FILING DATE: 1989-12-20
PRIOR APPLICATION NUMBER: US 07/487,732
PRIOR FILING DATE: 1990-03-02
PRIOR APPLICATION NUMBER: US 07/551,632
PRIOR FILING DATE: 1990-07-10
PRIOR APPLICATION NUMBER: US 07/379,778
PRIOR FILING DATE: 1989-07-14
PRIOR APPLICATION NUMBER: US 07/607,826
PRIOR FILING DATE: 1990-10-31
PRIOR APPLICATION NUMBER: US 07/595,468
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent in version 3.1
SEQ ID NO 10
LENGTH: 1017
TYPE: PRT
ORGANISM: Homo sapiens
US-08-468-996-10

Query Match 56.0%; Score 1955; DB 4; Length 1017;
Best Local Similarity 52.5%; Pred. No. 5.9e-116;
Matches 396; Conservative 38; Mismatches 182; Indels 138; Gaps 15;

Qy 1 GPPGPPGTLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60
Db 25 GNPGEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 84
Qy 61 -----GEAGLPKAGLGTGSP-----GSPGPP-----DGKTGPPGPA--- 90
Db 85 GVKHRRGVPGLDGAAGAGAPGVKGGSGPPGPPGPPGPPGPPGPPGPPG 144
Qy 91 GQDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 141
Db 145 GNDQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 204
Qy 142 -----GKDGEAGQGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 195
Db 205 GASGNPCTDGTIPGAKGSAGAPGAGAPGPPGPPGPPGPPGPPGPPG 264
Qy 196 GD-----LGAPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 237
Db 265 GEQPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 324
Qy 238 GPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 297
Db 325 GPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 384
Qy 298 GRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 357

[illegible]

RESULT 11

```

US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820

```

Query Match	55.7%	Score 1943;	DB 3;	Length 1060;
Best Local Similarity	52.1%	Pred. No. 3	5e-115;	
Matches 393;	Conservative 34;	Mismatches 189;	Indels 138;	Gaps 14;

QY	1	GPP	E	P	E	P	T	G	L	P	P	G	R	G	G	F	G	S	R	G	F	G	A	D	G	V	A	G	P	K	P	A	G	E	R	G	S	P	G	P	A	G	K	S	P	C	E	A	G	R	P	60											
Db	44	G	N	P	E	P	G	E	P	G	S	G	M	G	P	R	G	P	G	P	G	P	G	P	G	D	G	A	G	K	P	G	K	A	G	E	R	G	P	P	G	P	G	P	G	A	G	R	P	103													
QY	61	-----G																																E	A	G	L	P	G	A	K	L	T	G	S	P	---	G	S	P	G	---	D	K	T	G	P	P	G	P	A	---	90
Db	104	G	V	K	G	H	R	G	P	G	L	D	G	A	K	E	A	G	A	P	G	V	K	S	S	S	G	P	E	N	G	S	P	G	M	G	P	R	G	L	F	G	E	R	G	T	G	P	A	A	G	A	R	163									

QY	91	QDGRPGPPGPPGARGQAGVMGFP-----GPKGAAGEPGKAGRGVPPGPGVGA	141
Db	164	GNDGQPGAGPPGPPGVPAGGPGFCGAPGAKGEAGFTGARGPEGAQGRGEPGTGSGPGPA	223
QY	142	-----GKDGAGAGQPPGAPGAGERGEQGPAGSPGQGLPGPAGPPGGAAGKDEQGV	195
Db	224	GASGNPFGTDGIPGAKGSAGAPGIAGAGFPFGPRFPDQATGPIGPKQGTGKPGIAGFK	283
QY	196	GD---LGAPSGGAPAGRPPTGLP-----GPPRGCGPGSGRPPGADGVA	237
Db	284	GEQPKGPFPGAGPQAGPAGPAGEGKRGARGEPGVGFIGPPGERGAPGNRGFPQDGLA	343
QY	238	GPKPAGERSGFPAGPKSGPGEAGRPCEAGLPKAKGLTGPSGSPGPDGKTGPPGAGQD	297
Db	344	GPKAPGERGSGLAGPKGANGDPGRFGEPLGPGARGLTGRPGDAGPQGVKVGSGAPGED	403
QY	298	GRPGPPGPGARGQAGVMGPPGPKGAGRGPKKAGRGVPPGPGAVGAGKDGAGAGQPPP	357
Db	404	GRPFPPGQARGQDPVMGPPGPPGAGNGEPGKAGEKGLPGAPGLRGLPGKDGGTGAGPPP	463
QY	358	GPAGPAGERGEQGPAGSPFGQGLPGPAGPPGGAAGKPEQGVDPDLGAPGSPSGPAGE---	413
Db	464	GPAGPAGERGEQAGPSPGSGQLPGPPGPPGEGGKPDQGVPEACAPGLVGRGERGPP	523
QY	414	-----PGPTGLP-----GPPRGCGPSRGFPAGDVGAGPKGPA	447
Db	524	GERSSPGAQGLQGRGLPGTPTGTDGPKAGSGPAGFPGAQGPGLQGMPPGERGAAGTAGPK	583
QY	448	GERGS-----PGPAGPKGSPGEAGRPGEAGLPKAKGLTGS	483
Db	584	GDGRDVGEKGPEGAPKDGKGRGLTGIPGPPAGANGKEGVEGPPGPPAGSAGARGAPGER	643
QY	484	GSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEP-----GKA	534
Db	644	GETGPPGTSIAGTGPAGADGQPKAGKQGEAGQKDAGAFGPPQGPSCAPGPQGTGVTGPK	703
QY	535	GERGVGPPGCA-----VGPAGKDXGAGAGQPPGPA---GPAGERGEQGPAGSPQFQ	582
Db	704	GARGAQPPDGTGPPGAAGRVGPPGSGNGNPPGPPGPPSGKDGPKGARGDSGPPGRAGEP	763
QY	583	GLPGPAGPPGGAAGKPEQGVGPDLAGPAGSGPAG	616
Db	764	GLQGPAGPGEKGEPCDDGSPGAEGLPPGPOGLAG	797

RESULT 12

```

US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110889
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for
; TITLE OF INVENTION: in Body Fluid
; TITLE OF INVENTION: Method and U
; TITLE OF INVENTION: Disorders As
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patcin Release #1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436

```


PRIOR APPLICATION DATA: US/08/187,319
APPLICATION NUMBER: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match 55.7%; Score 1943; DB 3; Length 1418;

Best Local Similarity 52.1%; Pred. No. 4.5e-115;

Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGPTGLPGRGSGRFGPGADGAGVAGPKGAGRGSPGAGPKGSPGAGRP 60
DB 156 GNPGEFPGVSGPMPGPPGPKGDDGAGKPKGAGRGPPGQAGRGFTGCLP 215
QY 61 -----GEAGLPKAGLTGSP---GSPGP-----DGKTPGPPGA--- 90
DB 216 GVKHGRGPGGLDGAKEAGAPGVKSGSPGNGSPGPKGDDGAGKPKGAGRGPPGQAGRGFTGCLP 275
QY 91 GQGRGPPGPPGAGQAGVGMFP-----GPKGAAGPKGAGRGVGPAGVGA 141
DB 276 GNDGPGFAGPPGVPAGGPGFPAKAGKAGTGAAGKAGRGPPGQAGRGFTGCLP 335
QY 142 -----GKDEAGAAGPPGAGPAGGERGQGPAGSPGQGLPAGPPGAGKPKGQGV 195
DB 336 GASGNPGTDGIPGAGSAGAPGAGPAGPPGPPGQAGTGLPKGQTKGPKIAGFK 395
QY 196 GD---LGAPGSGPAGPPTGLP-----GPPGGRGSGRGPAGDVA 237
DB 396 GQGPAGPPGAGPQAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 455
QY 238 GPKGAGRGSPGAGPKGSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 297
DB 456 GPKGAGRGSPGAGPKGSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 515
QY 298 GRPGRGPPGAGQAGVGMFPGPKGAAGPKGAGRGVGPAGVGPAGKDEAGAGQGP 357
DB 516 GRPGRGPPGAGQAGVGMFPGPKGAAGPKGAGRGVGPAGVGPAGKDEAGAGQGP 575
QY 358 GPAGPAGRGSGPAGSPGQGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 413
DB 576 GPAGPAGRGSGPAGSPGQGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 635
QY 414 -----PGPTGLP-----GPPGGRGSGRGPAGDVA 447
DB 636 GERSGPAQGLGQGLPPTGTPGKAGSAGPAGPAGPAGPAGPAGPAGPAGPAG 695
QY 448 GERS-----PGPAGPKGSPGAGRGPEAGLPKAGLTGSP 483
DB 696 GDRDVGKPPGAGKDGKGLTGPFGPPGAGANGKGEVGPFPAGSAGAPGER 755
QY 484 GSPGPDGKTGPPGAGQDRPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 534
DB 756 GTGPPGTSGIAGPAGDQPCAKGEQGEAGQKAGPAGPAGPAGPAGPAGPAGPAG 815
QY 535 GERGVGPPGA-----VCPAGKDEAGAGQPPGA---GPAGRGEGQGPAGSPG 582

DB 816 GARGAQPPGATGPPGAAGRVGPPGSGNPGPPGPGSGKDGKPGKAGSDGPPGRAGER 875
QY 583 GLPAGPAGPAGKPKGEGQVPGDGLGAGPAGSPGAG 616
DB 876 GLQGPAGPAGKPKGEGQVPGDGLGAGPAGSPGAG 909

RESULT 13

US-09-010-999-1
Sequence 1, Application US/09010999
Patent No. 6132976
GENERAL INFORMATION:
APPLICANT: Poole, Anthony R.
APPLICANT: Hollander, Anthony P.
APPLICANT: Billingham, R. C.
TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,999
FILING DATE: 22-JAN-1998
CLASSIFICATION: 4335
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/448,501
FILING DATE: 17-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/984,123
FILING DATE: 04-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 032931/0212
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Human Type II Collagen
US-09-010-999-1

Query Match 55.7%; Score 1943; DB 3; Length 1418;

Best Local Similarity 52.1%; Pred. No. 4.5e-115;

Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGPTGLPGRGSGRFGPGADGAGVAGPKGAGRGSPGAGPKGSPGAGRP 60
DB 156 GNPGEFPGVSGPMPGPPGPKGDDGAGKPKGAGRGPPGQAGRGFTGCLP 215
QY 61 -----GEAGLPKAGLTGSP---GSPGP-----DGKTPGPPGA--- 90
DB 216 GVKHGRGPGGLDGAKEAGAPGVKSGSPGNGSPGPKGDDGAGKPKGAGRGPPGQAGRGFTGCLP 275
QY 91 GQGRGPPGPPGAGQAGVGMFP-----GPKGAAGPKGAGRGVGPAGVGA 141

QY 583 GLPGPAGPPGRAGKPEQGVPCDLGAPGSPGAG 616
DB 876 GLOGPAGPPGKGEFGDDGPGSAGPPGQGLAG 909

RESULT 15

US-09-570-573-20

; Sequence 20, Application US/09570573

; Patent No. 6342361

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/570,573

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/187,319

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Gogoris, Adda C

; REGISTRATION NUMBER: 29,714

; REFERENCE/DOCKET NUMBER: 4305/08701

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1418 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: COLLAGEN -ALPHA 1 (II)

; US-09-570-573-20

Query Match

Best Local Similarity 55.7%; Score 1943; DB 3; Length 1418;

Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGTGLPGPPGGRGGPSRFPAGADGVAGPKGAPAGERSGPPGAPGKSPGAEAGRP 60
DB 156 GNPGEFGEFVSGPMGRPPGPPGPKGDDGAGKPKGAGERPPGPPGQARGFPCTPLP 215
QY 61 -----GEAGLPCAKGLTGSP---GSPGP-----DGKTGPPGPA--- 90
DB 216 GVKHGRVPLDGAKEAGAPGVKSGSPGNGSPGPMGRPLGGERGRTGPGAGAR 275
QY 91 GQDGRPGPPGPPGARGQAGVMGFP-----GPKGAAGEPKKAGRGVPPGPPGAVGPA 141
DB 276 GNDGQPGPAGFPVGPAGGFPFGPAGKAGAGTARGPEGAQGPGEFTPGSPGPA 335
QY 142 -----GKDXGAGAGQPPGPPGAPGAGERGEGQGPAGSPGFGGLPGPAGPPGAEAKPGEGQGV 195

Search completed: June 17, 2005, 15:16:53

Job time : 31.1894 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 15:02:42 ; Search time 67.5856 Seconds
(without alignments)
3505.413 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGEPGTGLPGRPGRGG.....GEOGVGDLGAPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1714042 seqs, 36379560 residues

Total number of hits satisfying chosen parameters: 1714042

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3488	100.0	617	17	US-10-658-989A-2
2	3482	99.8	821	17	US-10-658-989A-3
3	2237.5	64.1	1014	17	US-10-901-816A-5
4	2237.5	64.1	1014	17	US-10-901-816A-6
5	2237.5	64.1	1014	17	US-10-901-816A-10
6	2237.5	64.1	1057	15	US-10-104-889-20
7	2237.5	64.1	1107	15	US-10-104-889-11
8	2237.5	64.1	1171	15	US-10-104-889-8
9	2237.5	64.1	1388	15	US-10-104-889-10
10	2237.5	64.1	1464	14	US-10-216-705-21
11	2237.5	64.1	1464	16	US-10-357-851-1

12	2237.5	64.1	1464	16	US-10-358-024-1	Sequence 1, Appli
13	2237.5	64.1	1464	16	US-10-788-792-150	Sequence 150, App
14	2235.5	64.1	1014	17	US-10-901-816A-9	Sequence 9, Appli
15	2235.5	64.1	1014	17	US-10-901-816A-11	Sequence 11, Appli
16	2234	64.0	1449	15	US-10-402-089-8	Sequence 8, Appli
17	2234	64.0	1449	15	US-10-402-072A-8	Sequence 8, Appli
18	2233.5	64.0	1461	16	US-10-468-091-25	Sequence 25, Appli
19	2233.5	64.0	1464	15	US-10-291-265-243	Sequence 243, App
20	2231.5	64.0	1057	15	US-10-104-889-16	Sequence 16, Appli
21	2228.5	63.9	1341	14	US-10-058-124-18	Sequence 18, Appli
22	2227.5	63.9	1464	10	US-09-918-715-261	Sequence 261, App
23	2227.5	63.9	1464	14	US-10-060-036-159	Sequence 159, App
24	2227.5	63.9	1464	14	US-10-171-311-36	Sequence 36, Appli
25	2227.5	63.9	1464	14	US-10-149-352-2	Sequence 2, Appli
26	2227.5	63.9	1464	14	US-10-177-293-65	Sequence 65, Appli
27	2227.5	63.9	1464	14	US-10-301-822-28	Sequence 28, Appli
28	2227.5	63.9	1464	16	US-10-734-564-79	Sequence 79, Appli
29	2227.5	63.9	1464	16	US-10-474-794-261	Sequence 261, App
30	2227.5	63.9	1464	16	US-10-723-860-2289	Sequence 2289, Ap
31	2227.5	63.9	1464	17	US-10-852-335A-157	Sequence 157, App
32	2225.5	63.8	1463	15	US-10-402-089-2	Sequence 2, Appli
33	2225.5	63.8	1463	15	US-10-402-072A-2	Sequence 2, Appli
34	2224.5	63.8	1014	17	US-10-901-816A-7	Sequence 7, Appli
35	2224.5	63.8	1014	17	US-10-901-816A-13	Sequence 13, Appli
36	2222.5	63.7	1014	17	US-10-901-816A-8	Sequence 8, Appli
37	2222.5	63.7	1169	15	US-10-104-889-6	Sequence 6, Appli
38	2205.5	63.2	1453	16	US-10-468-091-26	Sequence 26, Appli
39	2193.5	62.9	1014	17	US-10-901-816A-12	Sequence 12, Appli
40	2191	62.8	822	15	US-10-342-331-49	Sequence 49, Appli
41	1955	56.0	1017	16	US-10-639-286-10	Sequence 10, Appli
42	1946	55.8	1014	14	US-10-194-441A-1	Sequence 1, Appli
43	1943	55.7	1418	14	US-10-058-124-20	Sequence 20, Appli
44	1943	55.7	1418	16	US-10-468-091-5	Sequence 5, Appli
45	1938	55.6	544	17	US-10-658-989A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-658-989A-2
; Sequence 2, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-3
US-10-658-989A-2

Query Match	100.0%	Score 3488	DB 17	Length 617
Best Local Similarity	100.0%	Pred. No. 3.9e-167		
Matches 617	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	GPPGEPGTGLPGRPGRGGSGRFGADGAGKPGAGRGSGFGAGKPGSGRGP	60	
Db	1	GPPGEPGTGLPGRPGRGGSGRFGADGAGKPGAGRGSGFGAGKPGSGRGP	60	
Qy	61	GEAGLPGAKGLTSGPSGPPGDKTGPAGODRPPGPPGARGQGVGFGPKGAA	120	


```

Db      61  GEAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQAGVMGFGPPKGA 120
QY      121  GEPGKAGRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 180
Db      121  GEPGKAGRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 180
QY      181  GPPGAGKRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 240
Db      181  GPPGAGKRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 240
QY      241  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 300
Db      241  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 300
QY      301  GPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGVPPGPPGAGKDGAGAGQAGPPGPA 360
Db      301  GPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGVPPGPPGAGKDGAGAGQAGPPGPA 360
QY      361  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 420
Db      361  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 420
QY      421  GPPGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 480
Db      421  GPPGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 480
QY      481  GSPGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 540
Db      481  GSPGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 540
QY      541  GPPGAVGPPGAGKDGAGAGQAGPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 600
Db      541  GPPGAVGPPGAGKDGAGAGQAGPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 600
QY      601  GVPDGLGAPGSGPAG 617
Db      601  GVPDGLGAPGSGPAG 617

RESULT 2
US-10-658-989A-3
; Sequence 3, Application US/10658989A
; Publication No. US2005010151A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: compositions suitable for plasma substitution
; CURRENT APPLICATION NUMBER: US/10/658,989A
; PRIOR FILING DATE: 2003-09-10
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-4
US-10-658-989A-3

Query Match      99.8%; Score 3482; DB 17; Length 821;
Best Local Similarity 100.0%; Pred. No. 9.5e-167;
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GPPGEPGPTGLPFGPGERGSGRPPGAGDVGAGPKGPPAGERSGPPGPKGSGPCEAGRP 60
Db      1  GPPGEPGPTGLPFGPGERGSGRPPGAGDVGAGPKGPPAGERSGPPGPKGSGPCEAGRP 60
QY      61  GEAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQAGVMGFGPPKGA 120

```

```

Db      61  GEAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQAGVMGFGPPKGA 120
QY      121  GEPGKAGRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 180
Db      121  GEPGKAGRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 180
QY      181  GPPGAGKRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 240
Db      181  GPPGAGKRGVPPGPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGQAGSGPFGQGLPGPA 240
QY      241  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 300
Db      241  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 300
QY      301  GPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGVPPGPPGAGKDGAGAGQAGPPGPA 360
Db      301  GPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGVPPGPPGAGKDGAGAGQAGPPGPA 360
QY      361  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 420
Db      361  GPAGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 420
QY      421  GPPGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 480
Db      421  GPPGERSGPPGAGPKGSGPAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGR 480
QY      481  GSPGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 540
Db      481  GSPGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 540
QY      541  GPPGAVGPPGAGKDGAGAGQAGPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 600
Db      541  GPPGAVGPPGAGKDGAGAGQAGPPGPPGARGQAGVMGPPGKGAAGEPKKAGRGV 600
QY      601  GVPDGLGAPGSGPAG 616
Db      601  GVPDGLGAPGSGPAG 616

RESULT 3
US-10-901-816A-5
; Sequence 5, Application US/10901816A
; Publication No. US20050059703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: PP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 5
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-5

Query Match      64.1%; Score 2237.5; DB 17; Length 1014;
Best Local Similarity 60.2%; Pred. No. 1.7e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY      1  GPP-----GEPGPTGLPFGPGERGSGRPPGAGDVGAGPKGPPAGERSGPPGA 48
Db      280  GPPGPPAGEGKRGARGEPGPTGLPFGPGERGSGRPPGAGDVGAGPKGPPAGERSGPPGA 339
QY      49  GPKGSGPCEAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQA 108
Db      340  GPKGSGPCEAGRGAGLPGAKGLTGSFGSGPDPGKTPGPPGAGQDQGRPPGPPGARGQA 399

```



```

; ORGANISM: Homo sapiens
US-10-901-816A-10

Query Match      64.1%; Score 2237.5; DB 17; Length 1014;
Best Local Similarity 60.2%; Pred. No. 1.7e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GPGPTGLPPGPGGSGRFGPGADGVAGPKGAGRGSPGPA 48
DB 280 GPGFAGEGKRGARGGPTGLPPGPGGSGRFGPGADGVAGPKGAGRGSPGPA 339
QY 49 GPKSGPEAGRPAGLPGAKGLTSGSPGPPGKTPGPPAGQDGRPPGPPGARGQA 108
DB 340 GPKSGPEAGRPAGLPGAKGLTSGSPGPPGKTPGPPAGQDGRPPGPPGARGQA 399
QY 109 GVMGFPFGKGAAGPAGKAGRGVPGPGAVGAPAGKDGGAAGQPPGAPGARGGQGA 168
DB 400 GVMGFPFGKGAAGPAGKAGRGVPGPGAVGAPAGKDGGAAGQPPGAPGARGGQGA 459
QY 169 GSPGFQGLPGAPGPPGAGKAGRGVPGPGAVGAPAGKDGGAAGQPPGAPGARGGQGA 212
DB 460 GSPGFQGLPGAPGPPGAGKAGRGVPGPGAVGAPAGKDGGAAGQPPGAPGARGGQGA 519
QY 213 -----TGLPGPPGPGGSGRFGPGADGVAGPKGAGRGSPGPA 258
DB 520 GANGAPGNDGAKGDAGAPGAPGAGPGSQGAPGLQMPGERGAAGLPGKGDGADGSP 579
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPPGKTPGPPAGQDGRP 300
DB 580 GKDGVRLTGPFGPPGAPGAPGKSGSPGAPGTPGANGAPGCDRCEPFGPPGAPGPP 639
QY 301 GPPGPPGARGQ-----AGVMGFPFGKGAAGPAGKAGRGVPGPPGA----- 341
DB 640 GADGQPGAKGEPGDAGAKGDAGPPGAPGAPGPPGTPGPPGAPGAKGARGSPGATGPP 699
QY 342 -----VGPAGKDGGAAGQPPGAPG-----AGERGEGPAGSPGQGLPGAPGPPGAGKP 393
DB 700 GAAGRVGPPGSPGNAGPFPFGPPGAPGKGGKPRGETGAPRGPEVPPGPPGAPGKSGP 759
QY 394 GEGVPGDLGAPGSPGAPGPPGTPGLPPGPPGSGRFGPGADGVAGPKGAPGARGSP 453
DB 760 GADGAPAGAPGTPGFGIAGQGVVGLPGQGERGFPGLPGSPGEPKQGPSASGERGPP 819
QY 454 GPAGP---KGSPEAGRPGEAGLPGAKGLTSGSPGPPGKTPGPPAGQDGRPPGPP 510
DB 820 GPMGPPGLAGPPGESREGAPGAGSPGDRDGSFGAKGDRGETGAPGPPGAPGAPGPPV 879
QY 511 GARGQAGVMGFPFGKGAAGPAGKAGRGVPGPGAVGAPGAKDGE-----AGAQ 558
DB 880 GPAGKSGDRGETGAPGAPGAPGVPVGPAGARGPAGQGPGRGDKGTGEGQDGRGKHRGFSGLQ 939
QY 559 GPPGP-----AGPAGRGQGPAGSP---GFGQLPGAPGPPGAGKPGEGQGV 603
DB 940 GPPGPPGSPGEGQGPSAGSPGPPGPPGAGAPGKDGGLNGLPGLTGPFGPRGRTGDAGV 999
QY 604 GDLGAPGSPGAPG 616
DB 1000 GPPGPPGPPGPPG 1012

```

RESULT 6

```

US-10-104-889-20
; Sequence 20, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 22-Mar-2002
; APPLICATION NUMBER: US/10/104,889
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
; US-10-104-889-20

Query Match      64.1%; Score 2237.5; DB 15; Length 1057;
Best Local Similarity 60.2%; Pred. No. 1.7e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GPGPTGLPPGPGGSGRFGPGADGVAGPKGAPGARGSPGPA 48
DB 297 GPPGAPGEGKRGARGGFPGLPGPPGPGGSGRFGPGADGVAGPKGAPGARGSPGPA 356
QY 49 GPKSGPEAGRPAGLPGAKGLTSGSPGPPGKTPGPPAGQDGRPPGPPGARGQA 108
DB 357 GPKSGPEAGRPAGLPGAKGLTSGSPGPPGKTPGPPAGQDGRPPGPPGARGQA 416
QY 109 GVMGFPFGKGAAGPAGKAGRGVPGPGAVGAPGAKDGGAAGQPPGAPGARGGQGA 168
DB 417 GVMGFPFGKGAAGPAGKAGRGVPGPGAVGAPGAKDGGAAGQPPGAPGARGGQGA 476
QY 169 GSPGFQGLPGAPGPPGAGKAGRGVPGPGAVGAPGAKDGGAAGQPPGAPGARGGQGA 212
DB 477 GSPGFQGLPGAPGPPGAGKAGRGVPGPGAVGAPGAKDGGAAGQPPGAPGARGGQGA 536
QY 213 -----TGLPGPPGPGGSGRFGPGADGVAGPKGAGRGSPGPA 258
DB 537 GANGAPGNDGAKGDAGAPGAPGAGPGSQGAPGLQMPGERGAAGLPGKGDGADGSP 596
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPPGKTPGPPAGQDGRP 300
DB 597 GKDGVRLTGPFGPPGAPGAPGKSGSPGAPGTPGARGAPGDRGEPFGPPGAPGAP 656
QY 301 GPPGPPGARGQ-----AGVMGFPFGKGAAGPAGKAGRGVPGPPGA----- 341
DB 657 GADGQPGAKGEPGDAGAKGDAGPPGAPGAPGPPGTPGPNVGPAGKARGSPGATGFP 716
QY 342 -----VGPAGKDGGAAGQPPGAPG-----AGERGEGPAGSPGQGLPGAPGPPGAGKP 393
DB 717 GAAGRVGPPGSPGNAGPFPFGPPGAPGKGGKPRGETGAPRGPEVPPGPPGAPGKSGP 776
QY 394 GEGVPGDLGAPGSPGAPGPPGTPGLPPGPPGSGRFGPGADGVAGPKGAPGARGSP 453
DB 777 GADGAPAGAPGTPGFGIAGQGVVGLPGQGERGFPGLPGSPGEPKQGPSASGERGPP 836

```


QY 454 GPAGP---KCSGPEAGRPGEAGLPGAKGLTSGSPSPGDPGKTGTPPGAGQDGRPGPPG 510
Db 837 GPMGPPGLAGPBGSGREGAPGACGSGRDSGPKAGDRGETGAPGPGAFGAPG 896
QY 511 GARGOAGVMGPPGPKGAAGEGPKAGERGVPVPPGAVGAGPKDGE-----AGAQ 558
Db 897 GPAGSGDRGETGAPGAPGVPVPPGAGARGAPGQPRGDKGTGTGQDREGIKHGRGFSGLQ 956
QY 559 GPPGP-----AGPAGERGEOGPAGSP---GFQGLPGPAGPPGEAGKPGEOGVP 603
Db 957 GPPGPPGSGRGPBSGASGAPRGPPGSGAPGKDGGLNGLPGTGTGPPRGRGTGAGPV 1016
QY 604 GDLGAPGSPGAG 616
Db 1017 GPPGPPGPPGPPG 1029

RESULT 7
US-10-104-889-11
; Sequence 11, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1107 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-104-889-11

Query Match 64.1%; Score 2237.5; DB 15; Length 1107;
Best Local Similarity 60.2%; Pred.No. 1.8e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGTGLPGPGERGPGSGRFPFGADGVAGPKPGAGERGSPGA 48
Db 297 GPPGAGEGKRGAGEGPGTGLPGPGERGPGSGRFPFGADGVAGPKPGAGERGSPGA 356
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPSPGDPGKTGTPPGAGQDGRPGPPGARGQA 108
|||||

Db 357 GPKSGPGEAGRPGEAGLPGAKGLTSGSPSPGDPGKTGTPPGAGQDGRPGPPGARGQA 416
QY 109 GVMGFPGPKGAAAGEPGKAGERGVPVPPGAVGAGPKDGEAGAQGPAGPAGERGEGPA 168
Db 417 GVMGFPGPKGAAAGEPGKAGERGVPVPPGAVGAGPKDGEAGAQGPAGPAGERGEGPA 476
QY 169 GSPGFQGLPGPAGPPGEAGKPGEOGVPGDILGAPGSPGAGE-----PGP----- 212
Db 477 GSPGFQGLPGPAGPPGEAGKPGEOGVPGDILGAPGSPGARGERGPPGERGVQGPAGP 536
QY 213 -----TGLPFPGERGPGSGRGFPDADGVAGPKGAPGARGSGPAGPKGSP 258
Db 537 GANCAPGNDKAGDAGAPGAPGSGQAFGLQMGFERGCAAGLPGPKGDRGADGPKGADSP 596
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPSPGDPGKTGTPPGAGQDGRP 300
Db 597 GKDGVRLGTGPIGPPGAPGAGDKGESGSPGAPGTGARGAPGDRGEPGPPGAPGAGPP 656
QY 301 GPPGPPGARGQ-----AGVMGFPGPKGAAGEGPKAGERGVPVPPGA----- 341
Db 657 GADGQPKAGEPGDAGAKGDAGPPGAPGAPGPIGNVGAAPGAKGARGSGPPGATGFP 716
QY 342 -----VGPAGKDEAGAQGPAGP---AGERGEQGPAGSPGFGQLPGPAGPPGEAGKP 393
Db 717 GAAGRVGPPGSPGNAGPPGPPGAGKEGKGPRGETGAPRGVEVPPGPPGAGEKGP 776
QY 394 GEOGVPGDILGAPGSPGAPGPTGLPFPGERGPGSGRFPFGADGVAGPKGAPAGERGSP 453
Db 777 GADGAPAGPTGTPGQIAGQGVVGLPGQGERGFPGLPGPSGEPGKQGPSGASGERGPP 836
QY 454 GPAGP---KCSGPEAGRPGEAGLPGAKGLTSGSPSPGDPGKTGTPPGAGQDGRPGPP 510
Db 837 GPMGPPGLAGPBGSGREGAPGAEPSGRDGSFGAKGDRGETGAPGPGAPGAFGAPGV 896
QY 511 GARGOAGVMGPPGPKGAAGEGPKAGERGVPVPPGAVGAGPKDGE-----AGAQ 558
Db 897 GPAGSGDRGETGAPGAPGVPVPPGAPGARGPAGPQPRGDKGTGQDREGIKHGRGFSGLQ 956
QY 559 GPPGP-----AGPAGERGEOGPAGSP---GFQGLPGPAGPPGEAGKPGEOGVP 603
Db 957 GPPGPPGSGRGPBSGASGAPRGPPGSGAPGKDGGLNGLPGTGTGPPRGRGTGAGPV 1016
QY 604 GDLGAPGSPGAG 616
Db 1017 GPPGPPGPPGPPG 1029

RESULT 8
US-10-104-889-8
; Sequence 8, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889

FILING DATE: 22-Mar-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/169,768
 FILING DATE: 09-OCT-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: STEEN, JEFFREY S
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516) 228-8484
 TELEFAX: (516) 228-8516
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1171 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 US-10-104-889-8

Query Match 64.1%; Score 2237.5; DB 15; Length 1171;
 Best Local Similarity 60.2%; Pred. No. 1.8e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----GPPGTLGPPGGRGSGRFPAGDGVAGPKPAGERGSPGPA 48
 DB 297 GPPGAGEEGRGARGEPGTLGPPGGRGSGRFPAGDGVAGPKPAGERGSPGPA 356
 QY 49 GPKGSPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 357 GPKGSPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKGAAGHPKAGRGVPGPPGAVGPKDGEAGAGQPPGPPGAGERGQGA 168
 DB 417 GVMGFPKGAAGHPKAGRGVPGPPGAVGPKDGEAGAGQPPGPPGAGERGQGA 476
 QY 169 GSPGFQGLPGPAGPPEAGKPEQGVVGLGAPGSPGPA-----PGP----- 212
 DB 477 GSPGFQGLPGPAGPPEAGKPEQGVVGLGAPGSPGPA-----PGP----- 536
 QY 213 -----TGLPGPPGGRGSGRFPAGDGVAGPKPAGERGSPGPA 596
 DB 537 GANGAPNDGAKGDAGAPGAPGSGQAPGLQGMGREGAAGLPGPKDGRDAGPKGADGSP 596
 QY 259 GE-----ACRPFAGLPGAK-----GLTSGSPGPDGKTGPPGAGQDGRPP 300
 DB 597 GKDVRLGLTGTPTGPPGAGAPGDKSGSPGAGPTGARGAFGDRGEPGPPGAGFAGPP 656
 QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGHPKAGRGVPGPPGAVGPKDGEAGAGQPPGPPGAGERGQGA 341
 DB 657 GADGQFGAKGEPDAGAKGDAGPPGAGPAGPPGPIGVGAFKAGKAGSAGFPAGTATGPP 716
 QY 342 -----VGPAGKDEAGAGQPPGAGP-----AGRGSGGSPGSPGFGGLPGPAGPPGAGKP 393
 DB 717 GAAGRVGPPGPNAGPPGPPGAGKEGKGPRGETGAGRPGEVGPVPPGPPGAGEKSP 776
 QY 394 GEQGVPGDLGAPGSPGAPGPTGLPVPGRGSGRFPAGDGVAGPKPAGERGSP 453
 DB 777 GADGAPAGATTGPPQGIACQGVVGLGQGRGGRFPGLGSPGEPGKQPGSPGASGERGPP 836
 QY 454 GPAGP---KGSFGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGPP 510
 DB 837 GPMGPPGLAGPPGESREGAPGAGSPGFRDGSFAGKDRGETGAGPAGPPGAGPAGPVP 896
 QY 511 GARGQAGVMGFPKPKGAAGHPKAGRGVPGPPGAVGPKDGEAGAGQPPGPPGAGERGQGA 558
 DB 897 GPAGKSGDRGETGAPGAPGVPVGAARGPAGPQSPRGDKGETGECQDRGKIGHRGFSGLQ 956
 QY 559 GPPGP-----AGPACGERGQPPGAGSP-----GFGGLPGPAGPPGAGKPEQGV 603
 DB 957 GPPGPPGSPGEGSPGASGAPGPPGSPGAGAPGKDGGLNGLPGTIPGPPGRTGDAGPV 1016
 QY 604 GDLGAPGSPGAG 616

DB 1017 GPPGPPGPPGPPG 1029
 RESULT 9
 US-10-104-889-10
 ; Sequence 10, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRUSKIN, ELLIOT A.
 ; BUECHTER, DOUGLAS
 ; BROKAW, JANE
 ; ZHANG, GUANGHUI
 ; PAOLELLA, DAVID
 ; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DILWORTH & BARRESE
 ; STREET: 333 EARLE OVERTON BOULEVARD
 ; CITY: UNIONDALE
 ; STATE: NY
 ; COUNTRY: U.S.A.
 ; ZIP: 11553
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/104,889
 ; FILING DATE: 22-Mar-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/169,768
 ; FILING DATE: 09-OCT-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEEN, JEFFREY S
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 228-8484
 ; TELEFAX: (516) 228-8516
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1388 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-10-104-889-10

Query Match 64.1%; Score 2237.5; DB 15; Length 1388;
 Best Local Similarity 60.2%; Pred. No. 2.1e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----GPPGTLGPPGGRGSGRFPAGDGVAGPKPAGERGSPGPA 48
 DB 297 GPPGAGEEGRGARGEPGTLGPPGGRGSGRFPAGDGVAGPKPAGERGSPGPA 356
 QY 49 GPKGSPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 357 GPKGSPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKGAAGHPKAGRGVPGPPGAVGPKDGEAGAGQPPGPPGAGERGQGA 168
 DB 417 GVMGFPKGAAGHPKAGRGVPGPPGAVGPKDGEAGAGQPPGPPGAGERGQGA 476
 QY 169 GSPGFQGLPGPAGPPEAGKPEQGVVGLGAPGSPGPA-----PGP----- 212
 DB 477 GSPGFQGLPGPAGPPEAGKPEQGVVGLGAPGSPGPA-----PGP----- 536
 QY 213 -----TGLPGPPGGRGSGRFPAGDGVAGPKPAGERGSPGPA 596
 DB 537 GANGAPNDGAKGDAGAPGAPGSGQAPGLQGMGREGAAGLPGPKDGRDAGPKGADGSP 596

QY 259 GE-----AGPGEAGLPGAK-----GLTSPGSPGDKTGPAGQDGRP 300
 Db 597 GKGVRLTGTPIGPPGAPGAGDAGSGSPGAGTGAAGPAGDRGEPGPPGAPGAGP 656
 QY 301 GPPGPPGARGO-----AGVMGFPFGPKGAAGPFGKAGRGVPGPCA----- 341
 Db 657 GADQOPGAKGPPGADAGKADAGPPGAPGAPGPIGNVGAAGKAGSAGPAGATGFP 716
 QY 342 -----VGPAGKDGAGAGQPPGAPG-----AGERGEQGPAGSPGQGLPGPAGPGEAGKP 393
 Db 717 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGETGAGRGVEGPPGPPGAGKGP 776
 QY 394 GEQGVGDLGAPGSPGAPGPTGLPDPPERGSGRPPGADGVAGPKGAPAGERGSP 453
 Db 777 GADGPAGAGTTPGQIAGQGVVGLPGQORGERFGLPGSGFPGKQPSGASGERGPP 836
 QY 454 GPAGP---KGSPPGABRPGAGLPGAKGLTSPGSPGDKTGPAGQDGRP 510
 Db 837 GPMGPPGLAGPPGSGREGAGGAGSGPDRDGSAGKAGDRGTGAGPPGAGPAGAPGV 896
 QY 511 GARGQAVMGPPGPKGAAGEPKAGRGVPPGPAAGVAGPAGKDG-----AGAQ 558
 Db 897 GPAGKSGDRGTGAPGAPGVPAGARGPAGPQGPGRGDKGTGEGQDGRGKGRGFSGLQ 956
 QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFOGLPGPAGPGEAGKPGEQGV 603
 Db 957 GPPGPPGSGPQPSGASGAPGPPGSGAGAPGKQGLNGLPGTGPGRGRTGDAGPV 1016
 QY 604 GDLGAPGSPGAP 616
 Db 1017 GPPGPPGPPGPPG 1029

RESULT 10
 US-10-216-705-21
 ; Sequence 21, Application US/10216705
 ; Publication No. US2003009893A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Meristem Therapeutics, S.A.
 ; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
 ; TITLE OF INVENTION: obtaining Such and Their Uses
 ; FILE REFERENCE: 1149-3 DIV
 ; CURRENT APPLICATION NUMBER: US/10/216,705
 ; CURRENT FILING DATE: 2002-08-09
 ; PRIOR APPLICATION NUMBER: US 09/331,347
 ; PRIOR FILING DATE: 1999-08-17
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 21
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-216-705-21

Query Match 64.1%; Score 2237.5; DB 14; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 2.2e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPDPGERGSGRPPGADGVAGPKGAPAGERGSPGPA 48
 Db 458 GPPGAGEGKRGARGEFGPTGLPDPGERGSGRPPGADGVAGPKGAPAGERGSPGPA 517
 QY 49 GPKSGPAGRPGAGLPGAKGLTSPGSPGDKTGPAGQDGRP 108
 Db 518 GPKSGPAGRPGAGLPGAKGLTSPGSPGDKTGPAGQDGRP 577
 QY 109 GVMGPPGKGAAGEPKAGRGVPPGPAAGVAGPKDGAGAGQPPGAPAGERGEQGA 168
 Db 578 GVMGPPGKGAAGEPKAGRGVPPGPAAGVAGPKDGAGAGQPPGAPAGERGEQGA 637
 QY 169 GSPGFQGLPGAPGPPGAGKPGEQGVPDILGAPGSPGAGE-----PGP----- 212

Db 638 GSPGFQGLPGAPGPPGAGKPGEQGVPDILGAPGSPGARGRPPGPPGAPGPR 697
 QY 213 -----TGLPDPGERGSGRPPGADGVAGPKGAPAGERGSPGAPGKGP 258
 Db 698 GAGAPGNDGAKGADAGAPGAPGSGQACFLQMPGERGAAGLPGKGRDAGAPGADGSP 757
 QY 259 GE-----AGPGEAGLPGAK-----GLTSPGSPGDKTGPAGQDGRP 300
 Db 758 GKGVRLTGTPIGPPGAPGAGDAGSGSPGAGTGAAGPAGDRGEPGPPGAPGAGP 817
 QY 301 GPPGPPGARGO-----AGVMGFPFGPKGAAGPFGKAGRGVPGPCA----- 341
 Db 818 GADQOPGAKGPPGADAGKADAGPPGAPGAPGPIGNVGAAGKAGSAGPAGATGFP 877
 QY 342 -----VGPAGKDGAGAGQPPGAPG-----AGERGEQGPAGSPGQGLPGPAGPGEAGKP 393
 Db 878 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGETGAGRGVEGPPGPPGAGKGP 937
 QY 394 GEQGVGDLGAPGSPGAPGPTGLPDPPERGSGRPPGADGVAGPKGAPAGERGSP 453
 Db 938 GADGPAGAGTTPGQIAGQGVVGLPGQORGERFGLPGSGFPGKQPSGASGERGPP 997
 QY 454 GPAGP---KGSPPGABRPGAGLPGAKGLTSPGSPGDKTGPAGQDGRP 510
 Db 998 GPMGPPGLAGPPGSGREGAGGAGSGPDRDGSAGKAGDRGTGAGPPGAGPAGAPGV 1057
 QY 511 GARGQAVMGPPGPKGAAGEPKAGRGVPPGPAAGVAGPAGKDG-----AGAQ 558
 Db 1058 GPAGKSGDRGTGAPGAPGVPAGARGPAGPQGPGRGDKGTGEGQDGRGKGRGFSGLQ 1117
 QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFOGLPGPAGPGEAGKPGEQGV 603
 Db 1118 GPPGPPGSGPQPSGASGAPGPPGSGAGAPGKQGLNGLPGTGPGRGRTGDAGPV 1177
 QY 604 GDLGAPGSPGAP 616
 Db 1178 GPPGPPGPPGPPG 1190

RESULT 11
 US-10-357-851-1
 ; Sequence 1, Application US/10357851
 ; Publication No. US20040151731A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jicha, Douglas L.
 ; TITLE OF INVENTION: Method and Compositions Involving
 ; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
 ; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
 ; FILE REFERENCE: 13376US
 ; CURRENT APPLICATION NUMBER: US/10/357,851
 ; CURRENT FILING DATE: 2003-02-04
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo Sapiens
 US-10-357-851-1

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 2.2e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPDPGERGSGRPPGADGVAGPKGAPAGERGSPGPA 48
 Db 458 GPPGAGEGKRGARGEFGPTGLPDPGERGSGRPPGADGVAGPKGAPAGERGSPGPA 517
 QY 49 GPKSGPAGRPGAGLPGAKGLTSPGSPGDKTGPAGQDGRP 108
 Db 518 GPKSGPAGRPGAGLPGAKGLTSPGSPGDKTGPAGQDGRP 577
 QY 109 GVMGPPGKGAAGEPKAGRGVPPGPAAGVAGPKDGAGAGQPPGAPAGERGEQGA 168
 Db 518 GVMGPPGKGAAGEPKAGRGVPPGPAAGVAGPKDGAGAGQPPGAPAGERGEQGA 577
 QY 109 GVMGPPGKGAAGEPKAGRGVPPGPAAGVAGPKDGAGAGQPPGAPAGERGEQGA 168


```

Db 578 GVMGPPKGAAGEPKKAGRGVPPGAVGAVGAGKDGAGAGGPPGAGPAGERGEQCPA 637
QY 169 GSPFGQGLPGPAGPCEAGKPGEQGVPGDLGAPGSPGAGE-----PGP----- 212
Db 638 GSPFGQGLPGPAGPCEAGKPGEQGVPGDLGAPGSPGAGE-----PGP----- 697
QY 213 -----TGLPQPPGERGSGRSGRFPQADGVAGPKGAPGERSGSPGAPGKGP 258
Db 698 GANGAPGNDGAKGDAGAFGAGSQAGLQGMGERGAAGLPGPKGRDAGPKGADGSP 757
QY 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGDPGKTGTPGAPGQDGRP 300
Db 758 GKDQVRLGTGPIGPPGAGAGDKGSGSPGAGTGAAGAPGDRGPPGPPGAGP 817
QY 301 GPPGPPGARGO-----AGVMGPPGPKGAAGEPKKAGRGVPPGPPGA----- 341
Db 818 GADQPGKAGBPGDAGAKGDAGPPGAPGPPGPIGNVGAAGKAGSAGPPGATGFP 877
QY 342 -----VGPAGKDGAGAGGPPGAPG-----AGERGEQGPAGSPFGQGLPGPAGPPGAGKP 393
Db 878 GAAGRVGPPGSGNAGPPGPPGAGKGGKGRGETGAPRGVEVPPGPPGAGEKGP 937
QY 394 GEQGVPGDLGAPGSPGAGEPGTGLPQPPGERGSGRSGRFPQADGVAGPKGAPGERSG 453
Db 938 GADGPAGAPGTPGQIAGQGVVGLPQGRGERGFFGLPGSPGEPKQGPSSGASGERGP 997
QY 454 GPAGP-----KSPGAGRPGEAGLPGAKGLTSPGSPGDPGKTGTPGAPGQDGRP 510
Db 998 GPMGPPGLAGPPGSGREGAPGAGSGRDPGSPGAKGDRGETGAPGPPGAPGAPGVP 1057
QY 511 GARGQAGVMGPPGPKGAAGPFGKAGRGVPPGAVGAPGAKDGE-----AGAQ 558
Db 1058 GPAGKSGDRGETGAPGAPGVPVGPAGARGPAGPQGRDGETGEQDRIKGRHFGSLQ 1117
QY 559 GPPGP-----AGPAGERGQGPAGSP-----GFOGLPGPAGPPGAGKPGEQGV 603
Db 1118 GPPGPPSGEQGPGSAGSPGPRGPPGSAAGPKDGLNGLPGPIGPPGPRGRTGDAGVP 1177
QY 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 12
US-10-358-024-1
; Sequence 1, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Methods and Compositions Involving Blood
; FILE REFERENCE: T-Lymphocytes Reactivity with Collagen
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-358-024-1

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
Best Local Similarity 60.2%; Pred. No. 2.2e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGTPGLPQPPGERGSGRSGRFPQADGVAGPKGAPGERSGPPA 48
Db 458 GPPGAPGEGKRGAGEGPGTGLPQPPGERGSGRSGRFPQADGVAGPKGAPGERSGPPA 517
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSPGSPGDPGKTGTPGAPGQDGRPGRPPGARGQA 108

```

```

Db 518 GPKGSPGAGRPGEAGLPGAKGLTSPGSPGDPGKTGTPGAPGQDGRPGRPPGARGQA 577
QY 109 GVMGPPKGAAGEPKKAGRGVPPGAVGAVGAGKDGAGAGGPPGAPGAGERGEQCPA 168
Db 578 GVMGPPKGAAGEPKKAGRGVPPGAVGAVGAGKDGAGAGGPPGAPGAGERGEQCPA 637
QY 169 GSPFGQGLPGPAGPCEAGKPGEQGVPGDLGAPGSPGAGE-----PGP----- 212
Db 638 GSPFGQGLPGPAGPCEAGKPGEQGVPGDLGAPGSPGAGE-----PGP----- 697
QY 213 -----TGLPQPPGERGSGRSGRFPQADGVAGPKGAPGERSGSPGAPGKGP 258
Db 698 GANGAPGNDGAKGDAGAFGAGSQAGLQGMGERGAAGLPGPKGRDAGPKGADGSP 757
QY 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGDPGKTGTPGAPGQDGRP 300
Db 758 GKDQVRLGTGPIGPPGAGAGDKGSGSPGAGTGAAGAPGDRGPPGPPGAGP 817
QY 301 GPPGPPGARGO-----AGVMGPPGPKGAAGEPKKAGRGVPPGPPGA----- 341
Db 818 GADQPGKAGBPGDAGAKGDAGPPGAPGPPGPIGNVGAAGKAGSAGPPGATGFP 877
QY 342 -----VGPAGKDGAGAGGPPGAPG-----AGERGEQGPAGSPFGQGLPGPAGPPGAGKP 393
Db 878 GAAGRVGPPGSGNAGPPGPPGAGKGGKGRGETGAPRGVEVPPGPPGAGEKGP 937
QY 394 GEQGVPGDLGAPGSPGAGEPGTGLPQPPGERGSGRSGRFPQADGVAGPKGAPGERSG 453
Db 938 GADGPAGAPGTPGQIAGQGVVGLPQGRGERGFFGLPGSPGEPKQGPSSGASGERGP 997
QY 454 GPAGP-----KSPGAGRPGEAGLPGAKGLTSPGSPGDPGKTGTPGAPGQDGRP 510
Db 998 GPMGPPGLAGPPGSGREGAPGAGSGRDPGSPGAKGDRGETGAPGPPGAPGAPGVP 1057
QY 511 GARGQAGVMGPPGPKGAAGPFGKAGRGVPPGAVGAPGAKDGE-----AGAQ 558
Db 1058 GPAGKSGDRGETGAPGAPGVPVGPAGARGPAGPQGRDGETGEQDRIKGRHFGSLQ 1117
QY 559 GPPGP-----AGPAGERGQGPAGSP-----GFOGLPGPAGPPGAGKPGEQGV 603
Db 1118 GPPGPPSGEQGPGSAGSPGPRGPPGSAAGPKDGLNGLPGPIGPPGPRGRTGDAGVP 1177
QY 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 13
US-10-788-792-150
; Sequence 150, Application US/10788792
; Publication No. US20040191819A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; PRIOR FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 150
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-788-792-150

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
Best Local Similarity 60.2%; Pred. No. 2.2e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

```


QY 1 GPP-----GEPGPTGLPGRGCGSRGPGADGVAGPKGAGERSGPGA 48
Db |||||
QY 458 GPPGAGEGKRGARGEPGPTGLPGRGCGSRGPGADGVAGPKGAGERSGPGA 517
Db |||||
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPPGPPGARGQA 108
Db |||||
QY 518 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPPGPPGARGQA 577
Db |||||
QY 109 GVMGFPKGAAGPCKGAGRGVPGPGAVGPAKGDEAGAQGGPPGAGPAGERGQGA 168
Db |||||
QY 578 GVMGFPKGAAGPCKGAGRGVPGPGAVGPAKGDEAGAQGGPPGAGPAGERGQGA 637
Db |||||
QY 169 GSPGFGQLPGPAGPGRGAGKGEQGVPGDLGAPGSPGAG-----PGP----- 212
Db |||||
QY 638 GSPGFGQLPGPAGPGRGAGKGEQGVPGDLGAPGSPGAGRGPPGPPGAGPR 697
Db |||||
QY 213 -----TGLPFPGRGCGSRGPGADGVAGPKGAGERSGPGA 258
Db |||||
QY 698 GANCAPGNDGAKGDAGAPGAGPSQAGLQMPGCGERGAAGLPGPKGDRGADGPKGADGSP 757
Db |||||
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPGDKTGPAGQDGRP 300
Db |||||
QY 758 GKQGVRLGTGPIGPPGAGAPGDKGSGSPGAGTGCARGAPGDRGPPGPPGAGFAGPP 817
Db |||||
QY 301 GPPGPPGARGQ-----AGVMGFPCKGAAGEPKGAGRGVPGPGA----- 341
Db |||||
QY 818 GADQCPGAKGPGDAGAKGDAGPPGAPGAPGPPGPIGNVGAAPGAKGARGSAGPPGATGFP 877
Db |||||
QY 342 -----VGPAGKDEAGAQGGPPGAGP-----AGERGEOGPPGPGFQGLPGPAGPPGAGKP 393
Db |||||
QY 878 GAAGRVGPPGSGNAGPPGPPGAPGAKGKGKPRGPGTGPAGRGVEGPPGPPGAGEKGP 937
Db |||||
QY 394 GEQGVPGDLGAPGSPGAGEPGPTGLPFPGRGCGSRGPGADGVAGPKGAGERSG 453
Db |||||
QY 938 GADGAPGAGPTGPGQGIAGRGVVGLPGQRGERFPGLPGPSGPGKQSGSGASGERGPP 997
Db |||||
QY 454 GPAGP-----KSGPGEAGRPGEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPPGPP 510
Db |||||
QY 998 GPMGPPGLAGPPGSGREGAPGAGSGSPGDRGSPGAKGDRGTGAGPPGAPGAPGVP 1057
Db |||||
QY 511 GARGQAGVMGFPCKGAAGEPKGAGRGVPGPGAVGPAKGDE-----AGAQ 558
Db |||||
QY 1058 GPAGKSGDRGTGAPGAPGVPVGPAGAPGAPGQPRGDKGTGQDGRGKNGHGFSLQ 1117
Db |||||
QY 559 GPPGP-----AGPAGERGEOGPPGSP-----GFGQLPGPAGPPGAGEAGKPGEGVP 603
Db |||||
QY 1118 GPPGPPGSGPGQSGASGAPGPRPGSAGAPGKDGGLNGLPGPIGPPGPRGRTGDAGPV 1177
Db |||||

RESULT 14

US-10-901-816A-9

; Sequence 9, Application US/10901816A

; Publication No. US20050058703A1

; GENERAL INFORMATION:

; APPLICANT: Chang, Robert C.

; APPLICANT: Olsen, David R.

; APPLICANT: James, Polarek W.

; APPLICANT: Williams, Kim E.

; TITLE OF INVENTION: Gelatin Capsules

; FILE REFERENCE: FP0404 US

; CURRENT APPLICATION NUMBER: US/10/901,816A

; CURRENT FILING DATE: 2004-07-29

; PRIOR APPLICATION NUMBER: US 60/492,085

; PRIOR FILING DATE: 2003-08-01

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 9

; LENGTH: 1014

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-901-816A-9

Query Match 64.1%; Score 2235.5; DB 17; Length 1014;

Best Local Similarity 59.9%; Pred. No. 2.1e-104;

Matches 439; Conservative 24; Mismatches 153; Indels 117; Gaps 11;

QY 1 GPP-----GEPGPTGLPGRGCGSRGPGADGVAGPKGAGERSGPGA 48
Db |||||
QY 280 GPPGAGEGKRGARGEPGPTGLPGRGCGSRGPGADGVAGPKGAGERSGPGA 339
Db |||||
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPPGPPGARGQA 108
Db |||||
QY 340 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPPGPPGARGQA 399
Db |||||
QY 109 GVMGFPKGAAGPCKGAGRGVPGPGAVGPAKGDEAGAQGGPPGAGPAGERGQGA 168
Db |||||
QY 400 GVMGFPKGAAGPCKGAGRGVPGPGAVGPAKGDEAGAQGGPPGAGPAGERGQGA 459
Db |||||
QY 169 GSPGFGQLPGPAGPGRGAGKGEQGVPGDLGAPGSPGAG-----PGP----- 212
Db |||||
QY 460 GSPGFGQLPGPAGPGRGAGKGEQGVPGDLGAPGSPGAGRGPPGPPGAGPR 519
Db |||||
QY 213 -----TGLPFPGRGCGSRGPGADGVAGPKGAGERSGPGA 258
Db |||||
QY 520 GANCAPGNDGAKGDAGAPGAGPSQAGLQMPGCGERGAAGLPGPKGDRGADGPKGADGSP 579
Db |||||
QY 259 GE-----AGRPGEAGLPGAKGLTSGSPGPGDKTGPAGQDGRP 300
Db |||||
QY 580 GKQGVRLGTGPIGPPGAGAPGPPGSGSPGAGTGCARGAPGDRGPPGPPGAGFAGPP 639
Db |||||
QY 301 GPPGPPGARGQ-----AGVMGFPCKGAAGEPKGAGRGVPGPGA----- 341
Db |||||
QY 640 GADQCPGAKGPGDAGAKGDAGPPGAPGAPGPPGPIGNVGAAPGAKGARGSAGPPGATGFP 699
Db |||||
QY 342 -----VGPAGKDEAGAQGGPPGAGP-----AGERGEOGPPGPGFQGLPGPAGPPGAGKP 393
Db |||||
QY 700 GAAGRVGPPGSGNAGPPGPPGAPGAKGKGKPRGPGTGPAGRGVEGPPGPPGAGEKGP 759
Db |||||
QY 394 GEQGVPGDLGAPGSPGAGEPGPTGLPFPGRGCGSRGPGADGVAGPKGAGERSG 453
Db |||||
QY 760 GADGAPGAGPTGPGQGIAGRGVVGLPGQRGERFPGLPGPSGPGKQSGSGASGERGPP 819
Db |||||
QY 454 GPAGP-----KSGPGEAGRPGEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPPGPP 510
Db |||||
QY 820 GPMGPPGLAGPPGSGREGAPGAGSGSPGDRGSPGAKGDRGTGAGPPGAPGAPGVP 879
Db |||||
QY 511 GARGQAGVMGFPCKGAAGEPKGAGRGVPGPGAVGPAKGDE-----AGAQ 558
Db |||||
QY 880 GPAGKSGDRGTGAPGAPGVPVGPAGAPGAPGQPRGDKGTGQDGRGKNGHGFSLQ 939
Db |||||
QY 559 GPPGP-----AGPAGERGEOGPPGSP-----GFGQLPGPAGPPGAGEAGKPGEGVP 603
Db |||||
QY 940 GPPGPPGSGPGQSGASGAPGPRPGSAGAPGKDGGLNGLPGPIGPPGPRGRTGDAGPV 999
Db |||||
QY 604 GDLGAPGSPGAP 616
Db |||||
QY 1000 GPPGPPGPPGPPG 1012
Db |||||

RESULT 15

US-10-901-816A-11

; Sequence 11, Application US/10901816A

; Publication No. US20050058703A1

; GENERAL INFORMATION:

; APPLICANT: Chang, Robert C.

; APPLICANT: Olsen, David R.

; APPLICANT: James, Polarek W.

; APPLICANT: Williams, Kim E.

; TITLE OF INVENTION: Gelatin Capsules

; FILE REFERENCE: FP0404 US

; CURRENT APPLICATION NUMBER: US/10/901,816A

; CURRENT FILING DATE: 2004-07-29

; PRIOR APPLICATION NUMBER: US 60/492,085
 ; PRIOR FILING DATE: 2003-08-01
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 11
 ; LENGTH: 1014
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-901-816A-11

Query Match 64.1%; Score 2235.5; DB 17; Length 1014;
 Best Local Similarity 59.9%; Pred. No. 2.1e-104;
 Matches 439; Conservative 24; Mismatches 153; Indels 117; Gaps 11;

Qy	1	GPP-----GPPGTGLPPFGRGGSGRFPADGVAGPKGAGERGSPGPA	48
Db	280	GPPGAGEGKRARGEPGTGLFPFGRGGSGRFPADGVAGPKGAGERGSPGPA	339
Qy	49	GPKGSPGAGRPGEAGLPAGKLTGSPGSPDQKTPPGAGDGRPPGPPGARGQA	108
Db	340	GPKGSPGAGRPGEAGLPAGKLTGSPGSPDQKTPPGAGDGRPPGPPGARGQA	399
Qy	109	GVMGFPQKGAAGPPKAGRGVPPGAVGPKDCEAGAGQPPGAGPAGERGQGPA	168
Db	400	GVMGFPQKGAAGPPKAGRGVPPGAVGPKDCEAGAGQPPGAGPAGERGQGPA	459
Qy	169	GSPGQGLPGPAGPPGAGKFGQGVPGDLGAPGSPGAGE-----PGP----	212
Db	460	GSPGQGLPGPAGPPGAGKFGQGVPGDLGAPGSPGAGE-----PGP----	519
Qy	213	-----TGLPFPGRGGSGRFPADGVAGPKGAGERGSPGAGPR	519
Db	520	GANGAPNDGAKGDAGAPGAGPQGGAGLQCMFGERGAAGLPKPKGDRDAGPKGADGSP	579
Qy	259	GE-----AGRPGEAGLPAGKLTGSPGSPDQKTPPGAGQDGRP	300
Db	580	GKDGVRGLTGPIGPPGAGAPGPPGSGSPGAGTANGAPGDRGEPGPPGAGFAGPP	639
Qy	301	GPPGPPGARGQ-----AGVMGFPKGAAGEPKAGRGVPPGPA-----	341
Db	640	GADGQFQAGKGPDAGAKGDAGPPGAGPAGPPGPIGNVGAFCAGKAGSAGPPGATGPP	699
Qy	342	-----VGPAGKDGAGAGQPPGAGP---AGERGEQGPAGSPGQGLPGPAGPPGAGKP	393
Db	700	GAAGRVGPPGPPSGNAGPPGPPGAGKGGKPRGETGPAGRGVEVPPGPPGAGKSP	759
Qy	394	GEQGVPGDLGAPGSPGAGEPPTGLPFPGRGGSGRFPAGDGVAGPKGAGERGSP	453
Db	760	GADGPAGAPGTGPFQGIAGQGVVGLPQGRGGRFPGLFPGSGEPKQGPSGASGERGPP	819
Qy	454	GPAGP---KGSPEGAGRPGEAGLPAGKLTGSPGSPDQKTPPGAGQDGRPPGPP	510
Db	820	GPWGPGLAGPPGSGREGAPGAEPSGRDGFAGKDRGETGPAGPPGAPGAPGV	879
Qy	511	GARGQGVNMGFPKGAAGEPKAGRGVPPGAVGAGPKDGE-----AGAQ	558
Db	880	GPAGKSGDRGETGPAGPAGPVGAGARGPAGPQGRGDKGETGEQDGRGKRGHRCFSLQ	939
Qy	559	GPPGP-----AGPAGEREQGPAGSP---GFGGLPGPAGPPGEGAKGPGQGV	603
Db	940	GPPGPPGSGEQGPSGAGPAGPRPPGAGAPGKDLGLPGLPTGPPGPRGTGDAGPV	999
Qy	604	GDLGAPGSPGAP	616
Db	1000	GPPGPPGPPGPPG	1012

Search completed: June 17, 2005, 15:35:34
 Job time : 72.5856 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:42:55 ; Search time 21.1205 seconds
(without alignment)
2810.812 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGEPGPTGLPFPGERGG.....GEQGVGDLPAGPSGPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2228.5	63.9	1464	1 CGHUIS	collagen alpha 1(I)
2	2205.5	63.2	1453	2 S21626	collagen alpha 1(I)
3	2151.5	61.7	1042	1 CGHUIS	collagen alpha 1(I)
4	2011.5	57.7	671	1 CGRT1S	collagen alpha 1(I)
5	1944	55.7	1418	2 T45467	collagen alpha 1(I)
6	1943	55.7	1487	1 CGHU6C	collagen alpha 1(I)
7	1932	55.4	1419	2 A41182	collagen alpha 1(I)
8	1932	55.4	1487	2 B41182	collagen alpha 1(I)
9	1898	54.4	779	1 CGB01S	collagen alpha 1(I)
10	1890.5	54.2	1486	1 B40333	collagen alpha 1(I)
11	1883	54.0	1492	2 A40333	collagen alpha 1(I)
12	1857.5	53.3	1464	2 S59856	collagen alpha 1(I)
13	1856	53.2	1049	1 CGB07S	collagen alpha 1(I)
14	1853	53.1	1466	1 CGHU7L	collagen alpha 1(I)
15	1840	52.8	886	2 IS0694	collagen alpha 1(I)
16	1822.5	52.3	1496	1 CGHU2V	collagen alpha 2(V)
17	1804	51.7	1497	2 I49607	procollagen type V
18	1788.5	51.3	1373	1 A43291	collagen alpha 2(I)
19	1764	50.6	1366	1 CGHUIS	collagen alpha 2(I)
20	1702.5	48.8	1838	1 CGHU1V	collagen alpha 1(V)
21	1689.5	48.4	1843	2 S18803	collagen alpha 1(V)
22	1644	47.1	1806	1 CGHUIE	collagen alpha 1(X)
23	1640.5	47.0	1414	1 S23809	collagen alpha 2(I)
24	1612	46.2	1027	2 S28774	collagen alpha 2(I)
25	1587	45.5	1024	2 S18251	collagen alpha 1(X)
26	1585	45.4	1546	1 CGHU2E	collagen alpha 2(X)
27	1584	45.4	1691	1 S22917	collagen alpha 5(I)
28	1583.5	45.4	673	1 CGB06C	collagen alpha 1(I)
29	1568	45.0	2944	2 A54849	collagen alpha 1(V)

30 1556.5 44.6 1549 2 I48103
31 1550.5 44.5 888 2 S28791
32 1547 44.4 1690 1 CGHU1B
33 1530 43.9 1763 2 S16366
34 1526 43.8 964 1 CGCH2S
35 1522 43.6 1669 1 CGHU4B
36 1510 43.3 920 2 A45748
37 1508 43.2 1603 2 S23810
38 1503.5 43.1 1669 1 CGMS4B
39 1489.5 42.7 1670 1 CGHU3B
40 1481 42.5 1712 1 CGHU2B
41 1471.5 42.2 812 2 S31521
42 1463 41.9 1759 2 T29351
43 1459.5 41.8 1758 2 T29350
44 1456 41.7 1142 2 JX0369
45 1437.5 41.2 1752 2 A45407

ALIGNMENTS

RESULT 1

CGHUIS

collagen alpha 1(I) chain precursor - human

N:Alternate names: procollagen alpha 1(I) chain

C:Species: Homo sapiens (man)

C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 09-Jul-2004

C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11:

5269; A29439; I53466; A02852; I37247

R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.

Gene 67, 105-115, 1988

A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex

A:Reference number: I60114; MUID:88329734; PMID:2843432

A:Accession: I60114

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-369, 'L', 371-589 <DAL>

A:Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP

R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Procke

Biochem. J. 253, 919-922, 1988

A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t

A:Reference number: S01143; MUID:89025644; PMID:3178743

A:Accession: S01143

A:Molecule type: mRNA

A:Residues: 1-472 <TRO>

A:Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NID:

A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;

Nature 310, 337-340, 1984

A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of

A:Reference number: A93335; MUID:84270697; PMID:6462220

A:Accession: A93335

A:Molecule type: DNA

A:Residues: 1-58, 'Q', 60-181 <CHU>

A:Cross-references: EMBL:X00820; NID:G35857; PIDN:CAA25394.1; PID:G35658

R:Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.;

J. Biol. Chem. 263, 15151-15157, 1987

A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enha

A:Reference number: I55254; MUID:88033098; PMID:2822714

A:Accession: I55254

A:Status: translation not shown; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-45 <ROS>

A:Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388

R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A:Title: Regulatory elements in the first intron contribute to transcriptional control of

A:Reference number: A39943; MUID:88097389; PMID:3480516

A:Accession: A39943

A:Molecule type: DNA

A:Residues: 1-34 <BOR>

A:Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238

R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter structure and organization of the gene.
A:Reference number: 155237; MUID:85130970; PMID:2857713
A:Accession: 155237
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-34 <CH2>
A:Cross-references: GB:M10627; NID:G180383; PIDN:AAA51992.1; PID:G553226
R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, J. Biol. Chem. 265, 6312-6317, 1990
A:Title: In vivo and in vitro noncovalent association of excised alpha 1 (I) amino-terminal propeptides.
A:Reference number: A35233; MUID:90202908; PMID:2318855
A:Accession: A35233
A:Molecule type: protein
A:Residues: 33-52 <WR>
A:Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved R:Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F. EMBO J. 8, 1705-1710, 1989
A:Title: A base substitution in the exon of a collagen gene causes alternative splicing
A:Reference number: S09400; MUID:89356643; PMID:2767050
A:Accession: S09400
A:Molecule type: mRNA
A:Residues: 156-183 <WEI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) collagen gene.
A:Reference number: A90567; MUID:71038625; PMID:5529814
A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A:Accession: B90567
A:Molecule type: protein
A:Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z', 235, 'Z', 237, 'Z', 239, 'Z', 241, 'Z', 243, 'Z', 245, 'Z', 247, 'Z', 249, 'Z', 251, 'Z', 253, 'Z', 255, 'Z', 257, 'Z', 259, 'Z', 261, 'Z', 263, 'Z', 265, 'Z', 267, 'Z', 269, 'Z', 271, 'Z', 273, 'Z', 275, 'Z', 277, 'Z', 279, 'Z', 281, 'Z', 283, 'Z', 285, 'Z', 287, 'Z', 289, 'Z', 291, 'Z', 293, 'Z', 295, 'Z', 297, 'Z', 299, 'Z', 301, 'Z', 303, 'Z', 305, 'Z', 307, 'Z', 309, 'Z', 311, 'Z', 313, 'Z', 315, 'Z', 317, 'Z', 319, 'Z', 321, 'Z', 323, 'Z', 325, 'Z', 327, 'Z', 329, 'Z', 331, 'Z', 333, 'Z', 335, 'Z', 337, 'Z', 339, 'Z', 341, 'Z', 343, 'Z', 345, 'Z', 347, 'Z', 349, 'Z', 351, 'Z', 353, 'Z', 355, 'Z', 357, 'Z', 359, 'Z', 361, 'Z', 363, 'Z', 365, 'Z', 367, 'Z', 369, 'Z', 371, 'Z', 373, 'Z', 375, 'Z', 377, 'Z', 379, 'Z', 381, 'Z', 383, 'Z', 385, 'Z', 387, 'Z', 389, 'Z', 391, 'Z', 393, 'Z', 395, 'Z', 397, 'Z', 399, 'Z', 401, 'Z', 403, 'Z', 405, 'Z', 407, 'Z', 409, 'Z', 411, 'Z', 413, 'Z', 415, 'Z', 417, 'Z', 419, 'Z', 421, 'Z', 423, 'Z', 425, 'Z', 427, 'Z', 429, 'Z', 431, 'Z', 433, 'Z', 435, 'Z', 437, 'Z', 439, 'Z', 441, 'Z', 443, 'Z', 445, 'Z', 447, 'Z', 449, 'Z', 451, 'Z', 453, 'Z', 455, 'Z', 457, 'Z', 459, 'Z', 461, 'Z', 463, 'Z', 465, 'Z', 467, 'Z', 469, 'Z', 471, 'Z', 473, 'Z', 475, 'Z', 477, 'Z', 479, 'Z', 481, 'Z', 483, 'Z', 485, 'Z', 487, 'Z', 489, 'Z', 491, 'Z', 493, 'Z', 495, 'Z', 497, 'Z', 499, 'Z', 501, 'Z', 503, 'Z', 505, 'Z', 507, 'Z', 509, 'Z', 511, 'Z', 513, 'Z', 515, 'Z', 517, 'Z', 519, 'Z', 521, 'Z', 523, 'Z', 525, 'Z', 527, 'Z', 529, 'Z', 531, 'Z', 533, 'Z', 535, 'Z', 537, 'Z', 539, 'Z', 541, 'Z', 543, 'Z', 545, 'Z', 547, 'Z', 549, 'Z', 551, 'Z', 553, 'Z', 555, 'Z', 557, 'Z', 559, 'Z', 561, 'Z', 563, 'Z', 565, 'Z', 567, 'Z', 569, 'Z', 571, 'Z', 573, 'Z', 575, 'Z', 577, 'Z', 579, 'Z', 581, 'Z', 583, 'Z', 585, 'Z', 587, 'Z', 589, 'Z', 591, 'Z', 593, 'Z', 595, 'Z', 597, 'Z', 599, 'Z', 601, 'Z', 603, 'Z', 605, 'Z', 607, 'Z', 609, 'Z', 611, 'Z', 613, 'Z', 615, 'Z', 617, 'Z', 619, 'Z', 621, 'Z', 623, 'Z', 625, 'Z', 627, 'Z', 629, 'Z', 631, 'Z', 633, 'Z', 635, 'Z', 637, 'Z', 639, 'Z', 641, 'Z', 643, 'Z', 645, 'Z', 647, 'Z', 649, 'Z', 651, 'Z', 653, 'Z', 655, 'Z', 657, 'Z', 659, 'Z', 661, 'Z', 663, 'Z', 665, 'Z', 667, 'Z', 669, 'Z', 671, 'Z', 673, 'Z', 675, 'Z', 677, 'Z', 679, 'Z', 681, 'Z', 683, 'Z', 685, 'Z', 687, 'Z', 689, 'Z', 691, 'Z', 693, 'Z', 695, 'Z', 697, 'Z', 699, 'Z', 701, 'Z', 703, 'Z', 705, 'Z', 707, 'Z', 709, 'Z', 711, 'Z', 713, 'Z', 715, 'Z', 717, 'Z', 719, 'Z', 721, 'Z', 723, 'Z', 725, 'Z', 727, 'Z', 729, 'Z', 731, 'Z', 733, 'Z', 735, 'Z', 737, 'Z', 739, 'Z', 741, 'Z', 743, 'Z', 745, 'Z', 747, 'Z', 749, 'Z', 751, 'Z', 753, 'Z', 755, 'Z', 757, 'Z', 759, 'Z', 761, 'Z', 763, 'Z', 765, 'Z', 767, 'Z', 769, 'Z', 771, 'Z', 773, 'Z', 775, 'Z', 777, 'Z', 779, 'Z', 781, 'Z', 783, 'Z', 785, 'Z', 787, 'Z', 789, 'Z', 791, 'Z', 793, 'Z', 795, 'Z', 797, 'Z', 799, 'Z', 801, 'Z', 803, 'Z', 805, 'Z', 807, 'Z', 809, 'Z', 811, 'Z', 813, 'Z', 815, 'Z', 817, 'Z', 819, 'Z', 821, 'Z', 823, 'Z', 825, 'Z', 827, 'Z', 829, 'Z', 831, 'Z', 833, 'Z', 835, 'Z', 837, 'Z', 839, 'Z', 841, 'Z', 843, 'Z', 845, 'Z', 847, 'Z', 849, 'Z', 851, 'Z', 853, 'Z', 855, 'Z', 857, 'Z', 859, 'Z', 861, 'Z', 863, 'Z', 865, 'Z', 867, 'Z', 869, 'Z', 871, 'Z', 873, 'Z', 875, 'Z', 877, 'Z', 879, 'Z', 881, 'Z', 883, 'Z', 885, 'Z', 887, 'Z', 889, 'Z', 891, 'Z', 893, 'Z', 895, 'Z', 897, 'Z', 899, 'Z', 901, 'Z', 903, 'Z', 905, 'Z', 907, 'Z', 909, 'Z', 911, 'Z', 913, 'Z', 915, 'Z', 917, 'Z', 919, 'Z', 921, 'Z', 923, 'Z', 925, 'Z', 927, 'Z', 929, 'Z', 931, 'Z', 933, 'Z', 935, 'Z', 937, 'Z', 939, 'Z', 941, 'Z', 943, 'Z', 945, 'Z', 947, 'Z', 949, 'Z', 951, 'Z', 953, 'Z', 955, 'Z', 957, 'Z', 959, 'Z', 961, 'Z', 963, 'Z', 965, 'Z', 967, 'Z', 969, 'Z', 971, 'Z', 973, 'Z', 975, 'Z', 977, 'Z', 979, 'Z', 981, 'Z', 983, 'Z', 985, 'Z', 987, 'Z', 989, 'Z', 991, 'Z', 993, 'Z', 995, 'Z', 997, 'Z', 999, 'Z', 1001, 'Z', 1003, 'Z', 1005, 'Z', 1007, 'Z', 1009, 'Z', 1011, 'Z', 1013, 'Z', 1015, 'Z', 1017, 'Z', 1019, 'Z', 1021, 'Z', 1023, 'Z', 1025, 'Z', 1027, 'Z', 1029, 'Z', 1031, 'Z', 1033, 'Z', 1035, 'Z', 1037, 'Z', 1039, 'Z', 1041, 'Z', 1043, 'Z', 1045, 'Z', 1047, 'Z', 1049, 'Z', 1051, 'Z', 1053, 'Z', 1055, 'Z', 1057, 'Z', 1059, 'Z', 1061, 'Z', 1063, 'Z', 1065, 'Z', 1067, 'Z', 1069, 'Z', 1071, 'Z', 1073, 'Z', 1075, 'Z', 1077, 'Z', 1079, 'Z', 1081, 'Z', 1083, 'Z', 1085, 'Z', 1087, 'Z', 1089, 'Z', 1091, 'Z', 1093, 'Z', 1095, 'Z', 1097, 'Z', 1099, 'Z', 1101, 'Z', 1103, 'Z', 1105, 'Z', 1107, 'Z', 1109, 'Z', 1111, 'Z', 1113, 'Z', 1115, 'Z', 1117, 'Z', 1119, 'Z', 1121, 'Z', 1123, 'Z', 1125, 'Z', 1127, 'Z', 1129, 'Z', 1131, 'Z', 1133, 'Z', 1135, 'Z', 1137, 'Z', 1139, 'Z', 1141, 'Z', 1143, 'Z', 1145, 'Z', 1147, 'Z', 1149, 'Z', 1151, 'Z', 1153, 'Z', 1155, 'Z', 1157, 'Z', 1159, 'Z', 1161, 'Z', 1163, 'Z', 1165, 'Z', 1167, 'Z', 1169, 'Z', 1171, 'Z', 1173, 'Z', 1175, 'Z', 1177, 'Z', 1179, 'Z', 1181, 'Z', 1183, 'Z', 1185, 'Z', 1187, 'Z', 1189, 'Z', 1191, 'Z', 1193, 'Z', 1195, 'Z', 1197, 'Z', 1199, 'Z', 1201, 'Z', 1203, 'Z', 1205, 'Z', 1207, 'Z', 1209, 'Z', 1211, 'Z', 1213, 'Z', 1215, 'Z', 1217, 'Z', 1219, 'Z', 1221, 'Z', 1223, 'Z', 1225, 'Z', 1227, 'Z', 1229, 'Z', 1231, 'Z', 1233, 'Z', 1235, 'Z', 1237, 'Z', 1239, 'Z', 1241, 'Z', 1243, 'Z', 1245, 'Z', 1247, 'Z', 1249, 'Z', 1251, 'Z', 1253, 'Z', 1255, 'Z', 1257, 'Z', 1259, 'Z', 1261, 'Z', 1263, 'Z', 1265, 'Z', 1267, 'Z', 1269, 'Z', 1271, 'Z', 1273, 'Z', 1275, 'Z', 1277, 'Z', 1279, 'Z', 1281, 'Z', 1283, 'Z', 1285, 'Z', 1287, 'Z', 1289, 'Z', 1291, 'Z', 1293, 'Z', 1295, 'Z', 1297, 'Z', 1299, 'Z', 1301, 'Z', 1303, 'Z', 1305, 'Z', 1307, 'Z', 1309, 'Z', 1311, 'Z', 1313, 'Z', 1315, 'Z', 1317, 'Z', 1319, 'Z', 1321, 'Z', 1323, 'Z', 1325, 'Z', 1327, 'Z', 1329, 'Z', 1331, 'Z', 1333, 'Z', 1335, 'Z', 1337, 'Z', 1339, 'Z', 1341, 'Z', 1343, 'Z', 1345, 'Z', 1347, 'Z', 1349, 'Z', 1351, 'Z', 1353, 'Z', 1355, 'Z', 1357, 'Z', 1359, 'Z', 1361, 'Z', 1363, 'Z', 1365, 'Z', 1367, 'Z', 1369, 'Z', 1371, 'Z', 1373, 'Z', 1375, 'Z', 1377, 'Z', 1379, 'Z', 1381, 'Z', 1383, 'Z', 1385, 'Z', 1387, 'Z', 1389, 'Z', 1391, 'Z', 1393, 'Z', 1395, 'Z', 1397, 'Z', 1399, 'Z', 1401, 'Z', 1403, 'Z', 1405, 'Z', 1407, 'Z', 1409, 'Z', 1411, 'Z', 1413, 'Z', 1415, 'Z', 1417, 'Z', 1419, 'Z', 1421, 'Z', 1423, 'Z', 1425, 'Z', 1427, 'Z', 1429, 'Z', 1431, 'Z', 1433, 'Z', 1435, 'Z', 1437, 'Z', 1439, 'Z', 1441, 'Z', 1443, 'Z', 1445, 'Z', 1447, 'Z', 1449, 'Z', 1451, 'Z', 1453, 'Z', 1455, 'Z', 1457, 'Z', 1459, 'Z', 1461, 'Z', 1463, 'Z', 1465, 'Z', 1467, 'Z', 1469, 'Z', 1471, 'Z', 1473, 'Z', 1475, 'Z', 1477, 'Z', 1479, 'Z', 1481, 'Z', 1483, 'Z', 1485, 'Z', 1487, 'Z', 1489, 'Z', 1491, 'Z', 1493, 'Z', 1495, 'Z', 1497, 'Z', 1499, 'Z', 1501, 'Z', 1503, 'Z', 1505, 'Z', 1507, 'Z', 1509, 'Z', 1511, 'Z', 1513, 'Z', 1515, 'Z', 1517, 'Z', 1519, 'Z', 1521, 'Z', 1523, 'Z', 1525, 'Z', 1527, 'Z', 1529, 'Z', 1531, 'Z', 1533, 'Z', 1535, 'Z', 1537, 'Z', 1539, 'Z', 1541, 'Z', 1543, 'Z', 1545, 'Z', 1547, 'Z', 1549, 'Z', 1551, 'Z', 1553, 'Z', 1555, 'Z', 1557, 'Z', 1559, 'Z', 1561, 'Z', 1563, 'Z', 1565, 'Z', 1567, 'Z', 1569, 'Z', 1571, 'Z', 1573, 'Z', 1575, 'Z', 1577, 'Z', 1579, 'Z', 1581, 'Z', 1583, 'Z', 1585, 'Z', 1587, 'Z', 1589, 'Z', 1591, 'Z', 1593, 'Z', 1595, 'Z', 1597, 'Z', 1599, 'Z', 1601, 'Z', 1603, 'Z', 1605, 'Z', 1607, 'Z', 1609, 'Z', 1611, 'Z', 1613, 'Z', 1615, 'Z', 1617, 'Z', 1619, 'Z', 1621, 'Z', 1623, 'Z', 1625, 'Z', 1627, 'Z', 1629, 'Z', 1631, 'Z', 1633, 'Z', 1635, 'Z', 1637, 'Z', 1639, 'Z', 1641, 'Z', 1643, 'Z', 1645, 'Z', 1647, 'Z', 1649, 'Z', 1651, 'Z', 1653, 'Z', 1655, 'Z', 1657, 'Z', 1659, 'Z', 1661, 'Z', 1663, 'Z', 1665, 'Z', 1667, 'Z', 1669, 'Z', 1671, 'Z', 1673, 'Z', 1675, 'Z', 1677, 'Z', 1679, 'Z', 1681, 'Z', 1683, 'Z', 1685, 'Z', 1687, 'Z', 1689, 'Z', 1691, 'Z', 1693, 'Z', 1695, 'Z', 1697, 'Z', 1699, 'Z', 1701, 'Z', 1703, 'Z', 1705, 'Z', 1707, 'Z', 1709, 'Z', 1711, 'Z', 1713, 'Z', 1715, 'Z', 1717, 'Z', 1719, 'Z', 1721, 'Z', 1723, 'Z', 1725, 'Z', 1727, 'Z', 1729, 'Z', 1731, 'Z', 1733, 'Z', 1735, 'Z', 1737, 'Z', 1739, 'Z', 1741, 'Z', 1743, 'Z', 1745, 'Z', 1747, 'Z', 1749, 'Z', 1751, 'Z', 1753, 'Z', 1755, 'Z', 1757, 'Z', 1759, 'Z', 1761, 'Z', 1763, 'Z', 1765, 'Z', 1767, 'Z', 1769, 'Z', 1771, 'Z', 1773, 'Z', 1775, 'Z', 1777, 'Z', 1779, 'Z', 1781, 'Z', 1783, 'Z', 1785, 'Z', 1787, 'Z', 1789, 'Z', 1791, 'Z', 1793, 'Z', 1795, 'Z', 1797, 'Z', 1799, 'Z', 1801, 'Z', 1803, 'Z', 1805, 'Z', 1807, 'Z', 1809, 'Z', 1811, 'Z', 1813, 'Z', 1815, 'Z', 1817, 'Z', 1819, 'Z', 1821, 'Z', 1823, 'Z', 1825, 'Z', 1827, 'Z', 1829, 'Z', 1831, 'Z', 1833, 'Z', 1835, 'Z', 1837, 'Z', 1839, 'Z', 1841, 'Z', 1843, 'Z', 1845, 'Z', 1847, 'Z', 1849, 'Z', 1851, 'Z', 1853, 'Z', 1855, 'Z', 1857, 'Z', 1859, 'Z', 1861, 'Z', 1863, 'Z', 1865, 'Z', 1867, 'Z', 1869, 'Z', 1871, 'Z', 1873, 'Z', 1875, 'Z', 1877, 'Z', 1879, 'Z', 1881, 'Z', 1883, 'Z', 1885, 'Z', 1887, 'Z', 1889, 'Z', 1891, 'Z', 1893, 'Z', 1895, 'Z', 1897, 'Z', 1899, 'Z', 1901, 'Z', 1903, 'Z', 1905, 'Z', 1907, 'Z', 1909, 'Z', 1911, 'Z', 1913, 'Z', 1915, 'Z', 1917, 'Z', 1919, 'Z', 1921, 'Z', 1923, 'Z', 1925, 'Z', 1927, 'Z', 1929, 'Z', 1931, 'Z', 1933, 'Z', 1935, 'Z', 1937, 'Z', 1939, 'Z', 1941, 'Z', 1943, 'Z', 1945, 'Z', 1947, 'Z', 1949, 'Z', 1951, 'Z', 1953, 'Z', 1955, 'Z', 1957, 'Z', 1959, 'Z', 1961, 'Z', 1963, 'Z', 1965, 'Z', 1967, 'Z', 1969, 'Z', 1971, 'Z', 1973, 'Z', 1975, 'Z', 1977, 'Z', 1979, 'Z', 1981, 'Z', 1983, 'Z', 1985, 'Z', 1987, 'Z', 1989, 'Z', 1991, 'Z', 1993, 'Z', 1995, 'Z', 1997, 'Z', 1999, 'Z', 2001, 'Z', 2003, 'Z', 2005, 'Z', 2007, 'Z', 2009, 'Z', 2011, 'Z', 2013, 'Z', 2015, 'Z', 2017, 'Z', 2019, 'Z', 2021, 'Z', 2023, 'Z', 2025, 'Z', 2027, 'Z', 2029, 'Z', 2031, 'Z', 2033, 'Z', 2035, 'Z', 2037, 'Z', 2039, 'Z', 2041, 'Z', 2043, 'Z', 2045, 'Z', 2047, 'Z', 2049, 'Z', 2051, 'Z', 2053, 'Z', 2055, 'Z', 2057, 'Z', 2059, 'Z', 2061, 'Z', 2063, 'Z', 2065, 'Z', 2067, 'Z', 2069, 'Z', 2071, 'Z', 2073, 'Z', 2075, 'Z', 2077, 'Z', 2079, 'Z', 2081, 'Z', 2083, 'Z', 2085, 'Z', 2087, 'Z', 2089, 'Z', 2091, 'Z', 2093, 'Z', 2095, 'Z', 2097, 'Z', 2099, 'Z', 2101, 'Z', 2103, 'Z', 2105, 'Z', 2107, 'Z', 2109, 'Z', 2111, 'Z', 2113, 'Z', 2115, 'Z', 2117, 'Z', 2119, 'Z', 2121, 'Z', 2123, 'Z', 2125, 'Z', 2127, 'Z', 2129, 'Z', 2131, 'Z', 2133, 'Z', 2135, 'Z', 2137, 'Z', 2139, 'Z', 2141, 'Z', 2143, 'Z', 2145, 'Z', 2147, 'Z', 2149, 'Z', 2151, 'Z', 2153, 'Z', 2155, 'Z', 2157, 'Z', 2159, 'Z', 2161, 'Z', 2163, 'Z', 2165, 'Z', 2167, 'Z', 2169, 'Z', 2171, 'Z', 2173, 'Z', 2175, 'Z', 2177, 'Z', 2179, 'Z', 2181, 'Z', 2183, 'Z', 2185, 'Z', 2187, 'Z', 2189, 'Z', 2191, 'Z', 2193, 'Z', 2195, 'Z', 2197, 'Z', 2199, 'Z', 2201, 'Z', 2203, 'Z', 2205, 'Z', 2207, 'Z', 2209, 'Z', 2211, 'Z', 2213, 'Z', 2215, 'Z', 2217, 'Z', 2219, 'Z', 2221, 'Z', 2223, 'Z', 2225, 'Z', 2227, 'Z', 2229, 'Z', 2231, 'Z', 2233, 'Z', 2235, 'Z', 2237, 'Z', 2239, 'Z', 2241, 'Z', 2243, 'Z', 2245, 'Z', 2247, 'Z', 2249, 'Z', 2251, 'Z', 2253, 'Z', 2255, 'Z', 2257, 'Z', 2259, 'Z', 2261, 'Z', 2263, 'Z', 2265, 'Z', 2267, 'Z', 2269, 'Z', 2271, 'Z', 2273, 'Z', 2275, 'Z', 2277, 'Z', 2279, 'Z', 2281, 'Z', 2283, 'Z', 2285, 'Z', 2287, 'Z', 2289, 'Z', 2291, 'Z', 2293, 'Z', 2295, 'Z', 2297, 'Z', 2299, 'Z', 2301, 'Z', 2303, 'Z', 2305, 'Z', 2307, 'Z', 2309, 'Z', 2311, 'Z', 2313, 'Z', 2315, 'Z', 2317, 'Z', 2319, 'Z', 2321, 'Z', 2323, 'Z', 2325, 'Z', 2327, 'Z', 2329, 'Z', 2331, 'Z', 2333, 'Z', 2335, 'Z', 2337, 'Z', 2339, 'Z', 2341, 'Z', 2343, 'Z', 2345, 'Z', 2347, 'Z', 2349, 'Z', 2351, 'Z', 2353, 'Z', 2355, 'Z', 2357, 'Z', 2359, 'Z', 2361, 'Z', 2363, 'Z', 2365, 'Z', 2367, 'Z', 2369, 'Z', 2371, 'Z', 2373, 'Z', 2375, 'Z', 2377, 'Z', 2379, 'Z', 2381, 'Z', 2383, 'Z', 2385, 'Z', 2387, 'Z', 2389, 'Z', 2391, 'Z', 2393, 'Z', 2395, 'Z', 2397, 'Z', 2399, 'Z', 2401, 'Z', 2403, 'Z', 2405, 'Z', 2407, 'Z', 2409, 'Z', 2411, 'Z', 2413, 'Z', 2415, 'Z', 2417, 'Z', 2419, 'Z', 2421, 'Z', 2423, 'Z', 2425, 'Z', 2427, 'Z', 2429, 'Z', 2431, 'Z', 2433, 'Z', 2435, 'Z', 2437, 'Z', 2439, 'Z', 2441, 'Z', 2443, 'Z', 2445, 'Z', 2447, 'Z', 2449, 'Z', 2451, 'Z', 2453, 'Z', 2455, 'Z', 2457, 'Z', 2459, 'Z', 2461, 'Z', 2463, 'Z', 2465, 'Z', 2467, 'Z', 2469, 'Z', 2471, 'Z', 2473, 'Z', 2475, 'Z', 2477, 'Z', 2479, 'Z', 2481, 'Z', 2483, 'Z', 2485, 'Z', 2487, 'Z', 2489, 'Z', 2491, 'Z', 2493, 'Z', 2495, 'Z', 2497, 'Z', 2499, 'Z', 2501, 'Z', 2503, 'Z', 2505, 'Z', 2507, 'Z', 2509, 'Z', 2511, 'Z', 2513, 'Z', 2515, 'Z', 2517, 'Z', 2519, 'Z', 2521, 'Z', 2523, 'Z', 2525, 'Z', 2527, 'Z', 2529, 'Z', 2531, 'Z', 2533, 'Z', 2535, 'Z', 2537, 'Z', 2539, 'Z', 2541, 'Z', 2543, 'Z', 2545, 'Z', 2547, 'Z', 2549, 'Z', 2551, 'Z', 2553, 'Z', 2555, 'Z', 2557, 'Z', 2559, 'Z', 2561, 'Z', 2563, 'Z', 2565, 'Z', 2567, 'Z', 2569, 'Z', 2571, 'Z', 2573, 'Z', 2575, 'Z', 2577, 'Z', 2579, 'Z', 2581, 'Z', 2583, 'Z', 2585, 'Z', 2587, 'Z', 2589, 'Z', 2591, 'Z', 2593, 'Z', 2595, 'Z', 2597, 'Z', 2599, 'Z', 2601, 'Z', 2603, 'Z', 2605, 'Z', 2607, 'Z', 2609, 'Z', 2611, 'Z', 2613, 'Z', 2615, 'Z', 2617, 'Z', 2619, 'Z', 2621, 'Z', 2623, 'Z', 2625, 'Z', 2627, 'Z', 2629, 'Z', 2631, 'Z', 2633, 'Z', 2635, 'Z', 2637, 'Z', 2639, 'Z', 2641, 'Z', 2643, 'Z', 2645, 'Z', 2647, 'Z', 2649, 'Z', 2651, 'Z', 2653, 'Z', 2655, 'Z', 2657, 'Z', 2659, 'Z', 2661, 'Z', 2663, 'Z', 2665, 'Z', 2667, 'Z', 2669, 'Z', 2671, 'Z', 2673, 'Z', 2675, 'Z', 2677, 'Z', 2679, 'Z', 2681, 'Z', 2683, 'Z', 2685, 'Z', 2687, 'Z', 2689, 'Z', 2691, 'Z', 2693, 'Z', 2695, 'Z', 2697, 'Z', 2699, 'Z', 2701, 'Z', 2703, 'Z', 2705, 'Z', 2707, 'Z', 2709, 'Z', 2711, 'Z', 2713, 'Z', 2715, 'Z', 2717, 'Z', 2719, 'Z', 2721, 'Z', 2723, 'Z', 2725, 'Z', 2727, 'Z', 2729, 'Z', 2731, 'Z', 2733, 'Z', 2735, 'Z', 2737, 'Z', 2739, 'Z', 2741, 'Z', 2743, 'Z', 2745, 'Z', 2747, 'Z', 2749, 'Z', 2751, 'Z', 2753, 'Z', 2755, 'Z', 2757, 'Z', 2759, 'Z', 2761, 'Z', 2763, 'Z', 2765, 'Z', 2767, 'Z', 2769, 'Z', 2771, 'Z', 2773, 'Z', 2775, 'Z', 2777, 'Z', 2779, 'Z', 2781, 'Z', 2783, 'Z', 2785, 'Z', 2787, 'Z', 2789, 'Z', 2791, 'Z', 2793, 'Z', 2795, 'Z', 2797, 'Z', 2799, 'Z', 2801, 'Z', 2803, 'Z', 2805, 'Z', 2807, 'Z', 2809, 'Z', 2811, 'Z', 2813, 'Z', 2815, 'Z', 2817, 'Z', 2819, 'Z', 2821, 'Z', 2823, 'Z', 2825, 'Z', 2827, 'Z', 2829, 'Z', 2831, 'Z', 2833, 'Z', 2835, 'Z', 2837, 'Z', 2839, 'Z', 2841, 'Z', 2843, 'Z', 2845, 'Z', 2847, 'Z', 2849, 'Z', 2851, 'Z', 2853, 'Z', 2855

A:Status: translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1187-1194, 'C', 1196-1220 <CH>

A:Cross-references: GB:M23213; NID:9340842; PIDN:AA59363.1; PID:9499622

A:Note: mutant sequence from a patient with mild osteogenesis imperfecta

R:MacKellae, J.K.; Raasina, M.; Virta, A.; Vuorio, E.

Nucleic Acids Res. 16, 349, 1988

A:Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 63.9%; Score 2228.5; DB 1; Length 1464;

Best Local Similarity 60.0%; Pred. No. 4.2e-105;

Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

```
QY 1 GPP-----GERPCTGLPGRGSGRPFPGADGVAGPKPGAGERGSGCPA 48
Db |||
QY 458 GPPGPAGBEGKRGARFEPGLPFPGERGPGSRGFPFGADGVAGPKPGAGERGSGCPA 517
Db |||
QY 49 GPKSGPAGRPGAEAGLPCAKGLTSGSPGPDGKTPPGAGQDGRGPPGPPCARQQA 108
Db |||
QY 518 GPKSGPAGRPGAEAGLPCAKGLTSGSPGPDGKTPPGAGQDGRGPPGPPCARQQA 577
Db |||
QY 109 GVMGPPGPKGAAGEPGKAGRGVPPGAVGAGKDGAGAGQPPGPPAGPAGERGEGQGA 168
Db |||
QY 578 GVMGPPGPKGAAGEPGKAGRGVPPGAVGAGKDGAGAGQPPGPPAGPAGERGEGQGA 637
Db |||
QY 169 GSPGQGLPGPAGPGEAGKPGEGQVPGDLGAPGSGPAGE-----PGP----- 212
Db |||
QY 638 GSPGQGLPGPAGPGEAGKPGEGQVPGDLGAPGSGPARGRPFGRGVPQPPGPPGAPGR 697
Db |||
QY 213 -----TGLPQPCERGGSGSRGFPFGADGVAGPKPGAGERGSGCPAGPKGSP 258
Db |||
QY 698 GAGAPGNDGAKGDAGAPGAPGQAGLQGMFGERGAAGLPKPGKDRGDGPKGADGSP 757
Db |||
QY 259 GE-----AGRPGAEAGLPGAK-----GLTSGPSPGPDGKTPPGPAGQDGRP 300
Db |||
QY 758 GKGVGLTGTPIGPPGAGADKSGSPGAPGTGARGAPGDRGEPGPPGPPGAPGAGPP 817
Db |||
QY 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGEPGKAGRGVPGPPGA----- 341
Db |||
QY 818 GADGQPGAKGFGDAGAKGDAGPPGAPGAPGPPGPIGNVAPGAKGARGSGAPGATGPP 877
Db |||
QY 342 -----VGPAGKDGAEAGAGPPGAPGAP-----AGERGQPGAGSPFGLPGPAGPGEACKP 393
Db |||
QY 878 GAAGRVGPPGSGNAGPPGPPGAPGKGGKPRGETGAPRGVEGPPGPPGPPGAGEKSP 937
Db |||
QY 394 GEQGVGDLGAPGSPGAGEPGTGLPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 453
Db |||
QY 938 GADGAPAGPPTPGQIAGRGVGLPQORGERGPPGLPGSPGPKQSPGASGERGPP 997
Db |||
QY 454 GPAGP---KGSPGAEAGRPGAEAGLPGAKGLTSGSPGPDGKTPPGPAGQDGRGPPGPP 510
Db |||
QY 998 GPMGPPGLAGPPGSGREGAPCAEGSPGRDGSFGAKGDRGETGPPAGPPGAXGAGPAPV 1057
Db |||
QY 511 GARGQAGVMPGPPGKAGAECKAGRGVPPGPPGAVGAPGKDG-----AGAQ 558
Db |||
QY 1058 GPAGKSGDRGTGTPAGPAGVGPAGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1117
Db |||
QY 559 GPPG-----AGPAGERGEGQPGAPSP---GFQGLPGPAGPPGAEAGKPGEGQVP 603
Db |||
QY 1118 GPPGPPGSPGEGQPGSAGPAGPPGSPGAPGKDXGLNGLPGLPGLPGRGRTGDAGPV 1177
Db |||
QY 604 GDLGAPGSPGAPG 616
Db |||
QY 1178 GPPGPPGPPGPPG 1190
Db |||
```

RESULT 2

S21626

collagen alpha 1(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004

C:Accession: S57243; S16374; A23982; I49559; S39789; I48300; S21626

R:Li, S.W.; Khillan, J.; Prockop, D.J.

Matrix Biol. 14, 593-595, 1994

A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I

A:Reference number: S57243

A:Accession: S57243

A:Molecule type: mRNA

A:Residues: 1-1453 <IIS>

A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:9470673; PIDN:AAA8912.1; PID:9470673;

R:Metsaeranta, M.; Tonan, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16374

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1453 <MET>

A:Cross-references: EMBL:X57981; NID:950484; PIDN:CAA1046.1; PID:G50485

R:French, B.T.; Lee, W.H.; Maul, G.G.

Gene 39, 311-312, 1985

A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.

A:Reference number: A23982; MUID:86137403; PMID:3841523

A:Accession: A23982

A:Molecule type: mRNA

A:Residues: 518-1128 <FRE>

A:Cross-references: GB:M14423; NID:9192261; PIDN:AAA37333.1; PID:G192262

R:Monson, J.M.; Friedman, J.; McCarthy, B.J.

Mol. Cell. Biol. 2, 1362-1371, 1982

A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for

A:Reference number: I49559; MUID:83141374; PMID:6298597

A:Accession: I49559

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 735-1130 <RRS>

A:Cross-references: GB:M17491; NID:9192263; PIDN:AAA37334.1; PID:G192264

R:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.

Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984

A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads to

A:Reference number: I49557; MUID:84170331; PMID:6324198

A:Accession: I49557

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-25 <RE>

A:Cross-references: GB:K01688; NID:9192246; PIDN:AAA37330.1; PID:G553881

R:Penton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.P.

Biochim. Biophys. Acta 1216, 469-474, 1993

A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.

A:Reference number: S39789; MUID:94092741; PMID:8268229

A:Accession: S39789

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-13

R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.

Mol. Cell. Biol. 14, 5950-5960, 1994

A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect

A:Reference number: I48300; MUID:94344105; PMID:8065328

A:Accession: I48300

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

A:Cross-references: EMBL:X54876; NID:950486; PIDN:CAA38657.1; PID:G50487

C:Genetics:

A:Gene: COL1A1

A:introns: 770/3; 788/3; 806/3; 842/3; 860/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix;

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>

F:30-89/Domain: von Willebrand factor type C repeat homology <WVC>

F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>

F:1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match

Best Local Similarity 63.2%; Score 2205.5; DB 2; Length 1453;

Matches 432; Conservative 22; Mismatches 162; Indels 117; Gaps 11;

QY 1 GPP-----GEPPTGLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSPGA 48
 Db |||||
 447 GPPGAGEGKRGARFPGSLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSPGA 506
 QY 49 GPKGSPGAGRGEAGLCAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 Db |||||
 507 GPKGSPGAGRGEAGLCAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 566
 QY 109 GVMGPPGKGAAGFPGKAGRGVPPGAVGAPGAKDGEAGAQGPPGAPGAGERGQGA 168
 Db |||||
 567 GVMGPPGKGTAGFPGKAGRGVPPGAVGAPGAKDGEAGAQGPPGAPGAGERGQGA 626
 QY 169 GSPGQGLPGPAGPCEAGKFGEOGVPCDLGAPGSPGAGE-----PGP----- 212
 Db |||||
 627 GSPGQGLPGPAGPCEAGKFGEOGVPCDLGAPGSPGAGE-----PGP----- 686
 QY 213 -----TGLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSPGAGPR 258
 Db |||||
 687 GNNGAPGNDGAKGDTGAFGAFGSGAQGLQMPGEGAAGLPGPKGDRGDAGPKGADGSP 746
 QY 259 GE-----AGRPGEAGLCAKGLTSGPSGPD-----GKTGPPGAGQDGRP 300
 Db |||||
 747 GKDGAAGLGTGIPGPPGAGAPGDKGEAGSGPPGPTGARGAPGDRGEGAGPPGAGPP 806
 QY 301 GPPGPPGARGO-----AGVMGFPKPKAAGEPKKAGRGVPPGPA----- 341
 Db |||||
 807 GADQOPGAKGPPGDTGVKGDAAGPPGPPGAPGPPGPIGNVAPGPKGPRGAAGPPGATGFP 866
 QY 342 -----VGPAGKDGAGAQGPPGAGP-----AGERGEOGAGSPGQGLPGPAGPPGAGKP 993
 Db |||||
 867 GAAGRVGPPGSGNAGPPGPPGVGKGGKPRGETGPARPCEVPPGPPGAGKGGSP 926
 QY 394 GEQGVPCDLGAPGSPGAGPPGPTGLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSP 453
 Db |||||
 927 GADGAPGSPGTPGQGTAGQGVVGLPGQGRGERGFPGLPGSPGEPKQGFSGSGRGGPP 986
 QY 454 GPAGP-----KGSPEGAGRPGAGLPGAKGLTSGPSGPDGKTGPPGAGPPGAGKP 510
 Db |||||
 987 GPMGPPGLAGPPGSGREGSPGAGSGPRDCAKGLTSGPSGPDGKTGPPGAGPPGAGKP 1046
 QY 511 GARGQAGVMGFPKGAAGSPGKAGRGVPPG-----PGAV 546
 Db |||||
 1047 GPAGKNGDRGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1106
 QY 547 GPAGKNGDRGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 603
 Db |||||
 1107 GPPGSPGSPGEGQPSGASGAPGPPGPPGAGSGPKDGLNGLPGLPGPPGPRGRTGDSGA 1166
 QY 604 GDLGAPGSPGAPG 616
 Db |||||
 1167 GPPGPPGPPGPPG 1179

RESULT 3
 CGCH15
 collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
 C;Species: Gallus gallus (chicken)
 C;Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text_change 31-Mar-2000
 C;Accession: A90458; A90181; A02857
 R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
 Biochemistry 21, 2048-2055, 1982
 A;Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete prima
 A;Reference number: A90458; MUID:82231995; PMID:7093229
 A;Accession: A90458
 A;Molecule type: protein
 A;Residues: 1-1036 <HIG>
 A;Experimental source: skin
 A;Note: this is the latest in a series of papers from these workers elucidating the sequ
 R;Eyre, D.R.; Glimcher, M.J.
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972
 A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
 C;Species: Rattus norvegicus (Norway rat)
 A;Reference number: A90181; MUID:72243016; PMID:5047697
 A;Accession: A90181

A;Molecule type: protein
 A;Residues: 1037-1042 <EYR>
 A;Experimental source: skin
 A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
 C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
 C;Comment: Most of the prolines at the third position of the tripeptide repeating unit (C
 C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pos
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone carboxylic acid; trimer; t
 F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 61.7%; Score 2151.5; DB 1; Length 1042;
 Best Local Similarity 57.8%; Pred. No. 2.4e-101;
 Matches 424; Conservative 17; Mismatches 175; Indels 117; Gaps 10;

QY 1 GPP-----GEPPTGLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSPGA 48
 Db |||||
 296 GPPGAGEGKRGARFPGSLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSPGA 355
 QY 49 GPKGSPGAGRGEAGLCAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 Db |||||
 356 GPKGSPGAGRGEAGLCAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 415
 QY 109 GVMGPPGKGAAGFPGKAGRGVPPGAVGAPGAKDGEAGAQGPPGAPGAGERGQGA 168
 Db |||||
 416 GVMGPPGKGTAGFPGKAGRGVPPGAVGAPGAKDGEAGAQGPPGAPGAGERGQGA 475
 QY 169 GSPGQGLPGPAGPCEAGKFGEOGVPCDLGAPGSPGAGE-----PGP----- 212
 Db |||||
 476 GSPGQGLPGPAGPCEAGKFGEOGVPCDLGAPGSPGAGE-----PGP----- 535
 QY 213 -----TGLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSPGAGPR 251
 Db |||||
 536 GNNGAPGNDGAKGDTGAFGAFGSGAQGLQMPGEGAAGLPGPKGDRGDAGPKGADGSP 595
 QY 252 -----AGPKGSPGAGRP-----GEAGLPGAKGLTSGPSGPDGKTGPPGAGP----- 294
 Db |||||
 596 GKDGLRGLTGPFGPPGAGAPGDKGEAGPPGAPGTCARGAPGDRGEPGPPGAGFAGPP 655
 QY 295 -----QDGRPPGPPGAGRGVPPGAVGAPGAKDGEAGAQGPPGAPGAGERGSP 336
 Db |||||
 656 GADQOPGAKGPPGDTGVKGDAAGPPGPPGAPGPPGPIGNVAPGPKGPRGAAGPPGATGFP 715
 QY 337 GPPGAPGAGKDGAGAQGPPGPPGAGP-----RGEQGPAGSPGQGLPGPAGPPGAGKP 393
 Db |||||
 716 GAAGRVGPPGSGNIGLPGPPGAGKZGSKPRGETGPARPCEVPPGPPGAGKGGSP 775
 QY 394 GEQGVPCDLGAPGSPGAGPPGPTGLPGPPGGRGGSGRFGADGVAGPKGPPAGERGSP 453
 Db |||||
 776 GADGAPGSPGTPGQGTAGQGVVGLPGQGRGERGFPGLPGSPGEPKQGFSGSGRGGPP 835
 QY 454 GPAGP-----KGSPEGAGRPGAGLPGAKGLTSGPSGPDGKTGPPGAGPPGAGKP 510
 Db |||||
 836 GPMGPPGLAGPPGSGREGSPGAGSGPRDCAKGLTSGPSGPDGKTGPPGAGPPGAGKP 895
 QY 511 GARGQAGVMGFPKGAAGSPGKAGRGVPPG-----PGAV 558
 Db |||||
 896 GPAGKNGDRGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 955
 QY 559 GPP-----GPAGPAGRGEGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 603
 Db |||||
 956 GPPGPPGAGGEGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1015
 QY 604 GDLGAPGSPGAPG 616
 Db |||||
 1016 GPPGPPGPPGPPG 1028

RESULT 4
 CGRH15
 collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000

A;Residues: 419-567 <BU>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neutral salt-extracted rat skin collagen in neu
A;Reference number: A91209; MUID:74011954; PMID:4126850
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) in
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin
A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequen
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin colla
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status
F;9/Modified site: allysine (Lys) #status experimental
F;103,424,547/Binding site: carboxylate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 57.7%; Score 2011.5; DB 1; Length 671;
Best Local Similarity 59.5%; Pred. No. 1.8e-94;
Matches 396; Conservative 28; Mismatches 119; Indels 123; Gaps 14;

Qy	1	GPPEPTGLPGPGERGGPGS-----RGPRGA-----DGVAGPKPAGERGSPGPA	48
Db	29	GPAPGPGQGGPGEFGEFASGPMGRPPGPPGKNGDDGEAGKPRPQGRGPPGQ	88
Qy	49	GPKPGSGEAGRPGS-----AGLPKAKGLT-----GSPGSPGD---CKTGGPGGAGQD	93
Db	89	GARGLPETAGLPGMKGHGFSGLDGAKNTGAPGKPGSPGEGAPQMGPRGLPGR	148
Qy	94	GRPGPPPPGARGQGVNMGPPPKGAAGE-----PGKAGERGVPGPPGAVGAGKDEA	147
Db	149	GRPGPPGSAGARGDDGAVGAAGPPPGTGTGTPGPPGFAAGAKGEAGPQARGSEGPGVR	208
Qy	148	GAQPPPGPAGPAGERGEGP-----AGSPGFOGLPGPAGPPGEGAGKPEQ	192
Db	209	GEPPPPGAGAAGPAGNPGADGQPGAKGANGAPGTAGAPFPFGARPGSPQCPSPGAPGK	268
Qy	193	GVPCDLGAP-----GSPGA-----GEPPTGLPGPPGGRGGP	225
Db	269	GNSGEPAPGNKGTGAKGERPGPACVQPGPGPAGBEGKRGARGEPGPPSLGPPGGRGGP	328
Qy	226	GSRGFPGADGVAGPKPAGERGSPGAPGKSPGAGRPGEAGLPCAKGLTGS PGSPGD	285
Db	329	GSRGFPGADGVAGPKPAGERGSPGAPGKSPGAGRPGEAGLPCAKGLTGS PGSPGD	388
Qy	286	GKTGPPGADGDRPPPPGARGQAGVMGFPGBKGAAGERGPKAGERGVPPGAVGPA	345
Db	389	GKTGPPGACGZBGRPGPAGPFCARGQAGVMGFPBGKTAGGEPGKAGERGVPPGAVGPA	448
Qy	346	GKDEAGAQGPPGAPGAGERGEGPAGSPGFGPLPGPAGPPGEGAKPCEQGVPGDLGAP	405
Db	449	GKDEAGAQGAPGAPGAGERGEGPAGSPGFGPLPGPAGPPGEGAKPCEQGVPGDLGAP	508
Qy	406	GPSGPAGEPGTGLPGPPPPGRRGGPSRGFPFAGDGVAGPKPAGERGSPGAPGKSGEAG	465

D _b	509	GPGSGARGE						:	:	NGAPGBB	544
Q _y	466	GRGEAGLFCAKGLTCS	PSPGDPDKTTPPGAGDGRPGPPGGARGOAGVMPGPKR	---	RGFPERGVOQPPGPAGPRGN	---					
D _b	545	CAKGDGTGAPCAPGSQ	GAFLZGMSGLZGPPPPGS	---	PGZZGFSGASGAPGR	---					
Q _y	526	GAAGEFGKAGERGVCP	PGAVGPAKGACQAGAQPPGPPGAPAGERGEQGAPGSPQFGLP	---		---					
D _b	599	GSAGSFKGSLGLPCPI	IGPPFRGRTGBAGFSGFPFPGP	---		---					
Q _y	586	GPAGPPP	591								
D _b	641	GPPGPP	646								

RESULT 5

T45467 collagen alpha 1(II) chain precursor [imported] - horse
N;Alternate names: type II collagen
C;Species: Equus caballus (domestic horse)
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C;Accession: T45467
R;Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A;Description: Cloning of equine type II collagen and modulation of its expression in eq
A;Reference number: Z22977
A;Accession: T45467
A;Status: preliminary; translated from GB/EMBL/DBD
A;Molecule type: mRNA
A;Residues: 1-1418 <RIC>
A;Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:AB05773.1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match	55.7%;	Score 1944;	DB 2;	Length 1418;
Best Local Similarity	52.1%;	Pred. No. 7.3e-91;		
Matches	393;	Conservative 34;	Mismatches 189;	Indels 138; Gaps 14;
QY	1	GPPEPCPTGLPGPPDERGGPSGRPPGADGVAGPKGPAGERGSPGAPGKSPGEACRP	60	
Db	156	GNPGFGEFVSGPMGRPPGPPCKPGDGDGAGKPKSGERGPPGQCGARGFPGTGLP	215	
QY	61	-----GEAGLPGAKGLTGSPP-----GSPGP-----DKTQPPGPA---	90	
Db	216	GVKCHRGYPCLDGAKEAGAPGVKSGSGPENGSPGPMGRPLPCERGRTPGAGAAR	275	
QY	91	GDGRPPGPPGARGQAQVMGFPGPKGAAGFPKXAGER-----GVGPP-	134	
Db	276	GNDOGPGAGPPGVPAGGPPGAPCAKGAEGPTGARGPEGAQGRCEPTGSPGPA	335	
QY	135	-----PGAVGPAGKDGAEAGQPPGAPGAGERGEGP-----AGSPGFQGLP	177	
Db	336	GNAGNPTGDIPLGAKGSGAGAIAPFPGRGPPGPGATGLPGKGTGCEPCIAGFK	395	
QY	178	GPAGPPGAGKPGEGQVPGDLGAPFSPGAPGPGTGLPGPPGERGGPSGFFGADGVA	237	
Db	396	GEQGFKEFGPAGPQGAFGPAGECKRGARCEPGAGVPVPPGERKAPCNKGFPGQDGLA	455	
QY	238	GPKPAGERGSPGAPPKSGPGEAGRGPAEAGLPGAKGLTSGSPGPDGKTGPPGAGQD	297	
Db	456	GPKAGPERGSPGLAGPKGANGDPGRGEPGLPGARGLTGRPGDAGPQGVKVPSPGAGED	515	
QY	298	GRPPGPPGARGQAQVMGFPGPKGAAGEGPKAGERGVPVPGAVGPAKGKGEAGAGPP	357	
Db	516	GRPPGPPGQARGQFVGMFFPGPKANGECPKAGEKGLPGAPGLRGLPKGDGTGAAGPP	575	
QY	358	GPAPGAGERGQGPAGSPGQGLPCPAGPPGEAGKPGEQVPGDLGAPGPPSGPAGE---	414	
Db	576	GPAPGAGERGQGAPGSGFGQLPGPPGPPGEGCKPGDQVPGAGAPGLVGRPGRGFP	635	
QY	415	GPTGLP-----GPPERGGPSGRFPFADGVAGPKGPA	447	
Db	636	GERGSPGAQGLQARGHLPGTPTDGPKAGSPAGFPGAQPPGLOQMFGBERGAGIAGPK	695	

[illegible]

RESULT 6
CGHU6C

collagen alpha 1 (II) chain precursor [validated] - human
N;Alternate names: procollagen alpha 1 (II) chain
N;Contains: chondrocalcin; collagen alpha 1 (II) chain
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: A38513; S06715; S24770; A24828; S06496; A35428; A30147; A331116; S64674; S631
7250; I37253; I37253; I37254; I55338; I59535; I61910
R;Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A;Title: The human type II procollagen gene: identification of an additional protein-cod
A;Reference number: A38513; MUID:91184811; PMID:2081599
A;Accession: A38513
A;Molecule type: DNA
A;Residues: 1-103 <RYA>
A;Cross-references: UNIPROT:P02458; UNIPROT:Q14042; UNIPROT:Q16672; UNIPROT:Q12985; UNIP
R;Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A;Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A;Reference number: S06715; MUID:90067946; PMID:2587267
A;Accession: S06715
A;Molecule type: mRNA
A;Residues: 1-28, 'R', '99'-1487 <SU2>
A;Cross-references: EMBL:X16468; NID:929515; PIDN:CAA344488.1; PID:929516
A;Note: alternative splice form 1
R;Vikula, M.; Melsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A;Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A;Reference number: S24270; MUID:92344585; PMID:1637314
A;Accession: S24270
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-28 <VK>
A;Cross-references: EMBL:X58709; GB:S40537; NID:G35659
A;Note: this translation is not annotated in GenBank entry HSPROCE1, release 111.0
Gene 44, 11-15, 1986
A;Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A;Reference number: A24828; MUID:87031574; PMID:3021582
A;Accession: A24828
A;Molecule type: DNA
A;Residues: 1-8, 'T', 10-28 <NUN>
A;Cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237
R;Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A;Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1 (II) c
A;Reference number: S06496; MUID:90026318; PMID:2803268
A;Accession: S06496
A;Molecule type: mRNA
A;Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F',
A;Cross-references: EMBL:X16711; NID:G30040; PIDN:CAA34683.1; PID:G30041
A;Note: alternative splice form 1
R;Ryan, M.C.; Sandell, L.J.
Biol. Chem. 265, 10334-10339, 1990

A>Title: Differential expression of a cysteine-rich domain in the amino-terminal propeptide
 A:Reference number: A35428; MUID:90285153; PMID:2355003
 A:Accession: A35428
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 27-81,'L',83-103 <RYA2>
 A>Note: alternative splice form 2; splicing appears to be under developmental regulation
 R:Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
 Genomics 4, 438-441, 1989
 A>Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
 A:Reference number: A30147; MUID:89233138; PMID:2714801
 A:Accession: A30147
 A:Molecule type: DNA
 A:Residues: 104-157,'P',159-236 <SDM>
 A:Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25730; GB:M32168; GB
 R:Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
 Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
 A>Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
 A:Reference number: A94227; MUID:90370826; PMID:1975693
 A:Accession: A33116
 A:Molecule type: DNA
 A:Residues: 171-172,'C',174-175 <ALA>
 A>Note: mutant sequence from a family with primary generalized osteoarthritis
 R:Diab, M.; Wu, J.J.; Eyre, D.R.
 Biochem. J. 314, 327-332, 1996
 A>Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
 A:Reference number: S64673; MUID:96195147; PMID:8660302
 A:Accession: S64674
 A:Molecule type: protein
 A:Residues: 188-189,'X',191-195;1224-1230,'X',1232-1236 <DIA>
 R:Franc, S.; Marzin, E.; Bontillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
 Eur. J. Biochem. 234, 125-131, 1995
 A>Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil car
 A:Reference number: S63514; MUID:96096730; PMID:8529631
 A:Accession: S63514
 A:Molecule type: protein
 A:Residues: 243-261;575-590;756-763,'X',765-779 <FRA>
 R:Tiller, G.E.; Weis, M.A.; Polumbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
 Am. J. Hum. Genet. 56, 388-395, 1995
 A>Title: An RNA-splicing mutation (G+SVS20) in the type II collagen gene (COL2A1) in a
 A:Reference number: I38867; MUID:95150028; PMID:7847372
 A:Accession: I38867
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TIL1>
 A:Cross-references: EMBL:U15195; NID:9557053; PIDN:AAB60370.1; PID:9557054
 R:Ramirez, F.
 submitted to the EMBL Data Library, December 1988
 A:Reference number: S04892
 A:Accession: S04892
 A:Molecule type: mRNA
 A:Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'P',836-1214 <RAM>
 A:Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:g930050
 R:Vikkula, M.; Peltonen, L.
 FEBS Lett. 250, 171-174, 1989
 A>Title: Structural analyses of the polymorphic area in type II collagen gene.
 A:Reference number: S05000; MUID:89325561; PMID:2753125
 A:Accession: S05000
 A:Molecule type: DNA
 A:Residues: 630-640,'A',642-785 <VIK2>
 A:Cross-references: EMBL:X16158; NID:929951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
 PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
 R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
 J. Biol. Chem. 267, 22522-22526, 1992
 A>Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
 A:Reference number: A44309; MUID:93054548; PMID:1425602
 A:Accession: A44309
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA; mRNA
 A:Residues: 752-831,'PA',834,'P',836-1005,'K',1007-1036,'O',1038-1052,'E',1054-1068,'T',
 A:Cross-references: GB:L00977; NID:g180812; PIDN:AAB23314.1; PID:g258774
 A>Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence we
 A>Note: this translation is not annotated and this publication is not cited in GenBank e

A>Note: mutant sequence associated with perinatal lethal hypochondrogenesis
 R:Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
 A>Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
 A:Reference number: S16502; MUID:90251662; PMID:2339128
 A:Accession: S16502
 A:Molecule type: DNA
 A:Residues: 1164-1184,'GPSGKGANGIEGPT',1185-1199 <TIL2>
 A:Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
 A>Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
 R:Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
 Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
 A>Title: Identification and characterization of the human type II collagen gene (COL2A1)
 A:Reference number: A02858; MUID:85190534; PMID:3857598
 A:Accession: A02858
 A:Molecule type: DNA
 A:Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>
 A:Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
 R:Elima, K.; Vuorio, T.; Vuorio, E.
 Nucleic Acids Res. 15, 9499-9504, 1987
 A>Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) co
 A:Reference number: A27280; MUID:88067771; PMID:2825137
 A:Accession: A27280
 A:Molecule type: DNA; mRNA
 A:Residues: 1175-1487 <ELI>
 A:Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
 A:Experimental source: fetal epiphyseal cartilage
 R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
 Biochem. J. 237, 923-925, 1986
 A>Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
 A:Reference number: A57033; MUID:87099927; PMID:3800925
 A:Accession: A57033
 A:Molecule type: protein
 A:Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
 A>Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
 R:Strom, C.M.; Upholt, W.B.
 Nucleic Acids Res. 12, 1025-1038, 1984
 A>Title: Isolation and characterization of genomic clones corresponding to the human type
 A:Reference number: A21733; MUID:84118798; PMID:6320112
 A:Accession: A21733
 A:Molecule type: DNA
 A:Residues: 1245-1295 <STR1>
 A:Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
 A:Accession: B21733
 A:Molecule type: DNA
 A:Residues: 894-909,'PE', <STR2>
 A:Cross-references: GB:X01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
 R:Nunez, A.M.; Francmano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
 Biochemistry 24, 6343-6348, 1985
 A>Title: Isolation and partial characterization of genomic clones coding for a human pro
 gene.
 A:Reference number: A24561; MUID:86104139; PMID:3002437
 A:Accession: A24561
 A:Molecule type: DNA
 A:Residues: 1296-1358 <NUN2>
 A:Cross-references: GB:M12048; NID:g180017
 A>Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
 A>Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
 R:Sanigiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
 Nucleic Acids Res. 13, 2207-2225, 1985
 A>Title: Isolation and partial characterization of the entire human pro alpha 1(II) colla
 A:Reference number: I37249; MUID:85215609; PMID:2987845
 A:Accession: S59491
 A:Molecule type: DNA
 A:Residues: 7-28,'R',99-114;541-578;786-802;1055-1056,'N',1058-1068,'T',1070-1109;1200-1
 A:Accession: I84453
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 7-28 <SAN2>
 A:Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
 A>Note: the GenBank PID is based on an incorrect reading frame
 A:Accession: I37250
 A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 541-560 <SAN3>
A:CROSS-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A:Accession: I37251

```

Query Match      55.7%; Score 1943; DB 1; Length 1487;
Best Local Similarity 52.1%; Pred. No. 8.5e-91;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGTGLPDPGPGGPGSGFPCGADGAGVAGKPGAGRGSPGAGKPGSGFGEAGRP 60
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 GNPGEPPGVSQWPMGPRGPPGPKGDDGAGKPGKAGRGPPGQAGRGPGTGLP 284,
QY 61 -----GEAGLPGAKGLTGSP---GSPGP-----DGKTGPPGPA--- 90
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 GVKGHRGYPGLDGAKEAGAGVKGSGSPGSGPGRGLPGRGRTGAGAGAR 344
QY 91 GQDGRPGPPGPGARGQGVNGFP-----GPKAAGEPGKAGRGVPGPGVAGPA 141
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 GNDQPGGAGPVPVAGGPGFPCAFKAGAGTGTARGPGEAGQGPGRGPGTGGSPGA 404
QY 142 -----GKDGAGAGQGPAGPAGRGEGGPGAGSGRGARGEPGVPPIGPPGARGAGNRGPPGQGLA 524
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 GASNPGTGTGPKAGSAGAPGIAGAGTGTARGPGEAGTGTARGPGEAGTGTARGP 195
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 GD-----LGAPGSPGAGPPGTGLP-----GPPGGRGSGRSGFPFGADGVA 237
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
465 GEQPGKGPAGPQAGPAGPAGSGRGARGEPGVPPIGPPGARGAGNRGPPGQGLA 524
QY 238 GPKGPAGRGSPGAGPAGKPGSAGRPGEAGLPGKAGLTSRSPGDPKGTGPPGAGQD 297
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
525 GPKGAPGERSGLAGPKGANGPGRPEGLPGKAGLTSRSPGDPKGTGPPGAGQD 584
QY 298 GRPGRPPGARGQGVNGFPKGAAGEPGKAGRGVPGPGVAGVAGKAGAGQGP 357
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
585 GRPGRPPGARGQGVNGFPKGAAGEPGKAGRGVPGPGVAGVAGKAGAGQGP 644
QY 358 GPAGPAGRGEGQAGSPGQGLPGAGPAGKAGRGVPGPGVAGVAGKAGAGQGP 413
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 GPAGPAGRGEGQAGSPGQGLPGAGPAGKAGRGVPGPGVAGVAGKAGAGQGP 704
QY 414 -----PGPTGLP-----GPPGGRGSGRSGFPFGADGAGVAGKPA 447
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
705 GERGSPPAGQGLQFRLPGTGTGPKGASGAPGAGPGLQMPGREGAAGIAGPK 764
QY 448 GERGS-----PGAPGKSGPGEAGRGEGVPGPGAGSAGAGAPGER 824
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 GSPGDPGKTGPPGAGQDGRPPGPPGARGQGVNGFPKGAAGEP-----GKA 534
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
825 GETGPGTGTGAGPPGADQPGKAGRGVPGPGVAGVAGKAGAGQGP 884
QY 535 GERGVTPGPA-----VGPAGKDGAGAGQGPAGP---GPAGRGEGQAGSPGQF 582
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
885 GARGAQGPAGTGTGPAAGRVPPGSGNPGPPGPGSGKDGKPGKAGSDGPPGAGEP 944
QY 583 GLPGPAGPAGKAGRGVPGDLGAPGSPGAG 616
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
945 GLQGPAGPAGKAGRGVPGDLGAPGSPGAG 978

```

RESULT 7
A41182
collagen alpha 1(II) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A41182
R:Metaseraanta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: A41182

A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1419 <MET>
A:CROSS-references: GB:M65161
R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage
A:Reference number: A44885; MUID:91347939; PMID:1879363
A:Accession: A44885
A:Molecule type: DNA
A:Residues: 1-28 <CHE>
A:CROSS-references: GB:S63190; NID:g234368; PIDN:AAB19627.1; PID:g234369
A:Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:63192)
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer;
F:1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

```

Query Match      55.4%; Score 1932; DB 2; Length 1419;
Best Local Similarity 51.9%; Pred. No. 2.9e-90;
Matches 391; Conservative 36; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGTGLPDPGPGGPGSGFPCGADGAGVAGKPGAGRGSPGAGKPGSGFGEAGRP 60
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 GNPGEPPGVSQWPMGPRGPPGPKGDDGAGKPGKAGRGPPGQAGRGPGTGLP 216
QY 61 -----GEAGLPGAKGLTGSP---GSPGP-----DGKTGPPGPA--- 90
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 GVKGHRGYPGLDGAKEAGAGVKGSGSPGSGPGRGLPGRGRTGAGAGAR 276
QY 91 GQDGRPGPPGPGARGQGVNGFP-----GPKAAGEPGKAGRGVPGPGVAGPA 141
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
277 GNDQPGGAGPVPVAGGPGFPCAFKAGAGTGTARGPGEAGTGTARGPGEAGTGTARGP 336
QY 142 GKDGE-----AGAOGPPGAGPAGRGEGQGP-----AGSPFGQLP 177
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 GASNPGTGTGPKAGSAGAPGIAGAGTGTARGPGEAGTGTARGPGEAGTGTARGP 396
QY 178 GPAGPAGRGSPGAGPAGKPGSAGRPGEAGLPGKAGLTSRSPGDPKGTGPPGAGQD 297
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
457 GPKGAPGERSGLAGPKGANGPGRPEGLPGKAGLTSRSPGDPKGTGPPGAGQD 516
QY 298 GRPGRPPGARGQGVNGFPKGAAGEPGKAGRGVPGPGVAGVAGKAGAGQGP 357
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
517 GRPGRPPGARGQGVNGFPKGAAGEPGKAGRGVPGPGVAGVAGKAGAGQGP 576
QY 358 GPAGPAGRGEGQAGSPGQGLPGAGPAGKAGRGVPGPGVAGVAGKAGAGQGP 413
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
637 GERGSPPAGQGLQFRLPGTGTGPKGASGAPGAGPGLQMPGREGAAGIAGPK 696
QY 448 GERGSPPAGKAGRGVPGDLGAPGSPGAG 489
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
697 GDRGDPGKAGRGVPGDLGAPGSPGAG 756
QY 490 GKTGPPGAGQDGRPPGPPGAG-----GQAGVMGFPKGAAGEP-----GKA 534
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 GETGPPGAGTGTGPAAGRVPPGSGNPGPPGPGSGKDGKPGKAGSDGPPGAGEP 816
QY 535 GERGVTPGPA-----VGPAGKDGAGAGQGPAGP---GPAGRGEGQAGSPGQF 582
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
817 GARGAQGPAGTGTGPAAGRVPPGSGNPGPPGPGSGKDGKPGKAGSDGPPGAGEP 876
QY 583 GLPGPAGPAGKAGRGVPGDLGAPGSPGAG 616
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
877 GLQGPAGPAGKAGRGVPGDLGAPGSPGAG 910

```


RESULT 8

B41182 collagen alpha 1(I) chain precursor (long splice form) - mouse
C:Species: Mus musculus (house mouse)
C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 09-Jul-2004
C:Accession: B41182
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A>Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: B41182
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1487 <NET>
A:Cross-references: UNIPROT:Q62031; UNIPROT:Q62032; UNIPROT:Q62033; GB:I65161
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
F:33-91/Domain: von Willebrand factor type C repeat homology <WVC>
F:1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 55.4%; Score 1932; DB 2; Length 1487;
Best Local Similarity 51.9%; Pred. No. 3e-90;
Matches 391; Conservative 36; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPCEPGPTGLPPGGERGGPGRGFPAGDVGAPKGPAGERSPGAGPKGPSGAGR 60
DB :|||:|||||:
225 GNPCPEGVSGPMGRPDPAGKPKGDGEAKPKSGERGLPGPQGARGPFGTGLP 284
QY 61 -----GEAGLPKAGLGTSP---GSRCP-----DGKTGPPGA--- 90
DB :|||:|||||:
285 GVKGHRYPGLDGAKGAGAPGVKGSPEGNSGPMGPRGLPGBERTGPAGAAGAR 344
QY 91 QDQRCPGPPPPGARGQAQVGMGFF-----GPKGAAGEPKAGRGVPGVPAGVGA 141
DB :|||:|||||:
345 GNDQCFPAGPPGVGPPGPFPGAPCAKGEAGPTGARPEGAQSRRGFNPSPGPA 404
QY 142 GKDE-----AGAQPPGPAGPAGERGEQQP-----AGSPFGQLP 177
DB :|||:|||||:
405 GASGNPCTDTPGAKGSAGAPGIAGAPFPGRGPPPOGATGPLGPKQAQEPGIAGFK 464
QY 178 GPAGPPAEAGKPGEQGVPDLDGAPPSGAPAGEPTGLPGRPERGCGPSRGFPADVA 237
DB :|||:|||||:
465 GDQCFKETGPAGPQAGPAGPAGEKKRGAERGEPCGAGPIGPPGERGAPNRGFPQDGLA 524
QY 238 GPKGPAGERGSPGAPKGPSBAGRPGEAGLPGKAGLTGSPSGPSGPDGKTGPPGAGOD 297
DB :|||:|||||:
525 GPKGAPERSPGLAGPKGANGDPGRCEEGLPGARGLTRPGDAGPQGVGSGAPGED 584
QY 298 GRPGPPPPGARGQAQVMGFPFGPKAAAGEPKKAGERGVPVPGAVGPAGKDXGAGAQGP 357
DB :|||:|||||:
585 GRPGPPPQCARQPQVGMGFPFGPKGANPEPKAGEKLAGAPGLRGLPKDGETGAAGPP 644
QY 358 GPAPAGERGEQAPGSPGFGQLPGPAPGPGEAGKPGEQGVVDLDGHAPPSPGAGE---- 413
DB :|||:|||||:
645 GPSGPAGERGEQAPGSPGFGQLPGPFPFGEEGKQDQGITGEAGAPGLVGPGRGERGFP 704
QY 414 -----PGTGLP-----GPPCERGGPSRGPAGDGVAGPKGPA 447
DB :|||:|||||:
705 GERGSPAQGLQGRPLGFTGTDTGPKAAAGPDGPPGAQGGPLQGMPSRGAAGIAGPK 764
QY 448 GERGSPGAPGKSGPGE-----AGRPGEAGLPKAGLGTGSPSGPGPD 489
DB :|||:|||||:
765 GDRCDVKEKGEPAGPKDGGRLTGRTIGPPGAPAGNKEKEVEVPPGSGSTGARGAPEP 824
QY 490 KTKTPPGPADGDRGPPPPGPAR-----GQAGVMGFPGPKGAAGBP-----GKA 534
DB :|||:|||||:
825 GETGPPGAPGAPPGADGQPGAKDQGEAGQKDAGAPGQGPSGAPGPGQTGVTGPK 884
QY 535 GERGVPPPGA-----VGRPAGKDEAGAQGPFGPA---GPAGERGQGPAGSPFGQ 582
DB :|||:|||||:
885 GARGAQPPGATGTFPGAAGRVPFGANNGNFPAGPPGAPGKDGPKGVGDSGPPGRAGDP 944

Qy 583 GLPGAPGPPGAGKGGEGGVGDGLGAPGPSAG 616
|| |||| |||| |||| |||| |||| |||| |||| ||||
Db 945 GLQGPAPGAPGKGGEGGVGDGLGAPGPSAG 978

RESULT 9

COLLAGS
collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)
C/Species: Bos primigenius taurus (cattle)
C/Date: 24-Apr-1994 #sequence revision 31-Dec-1993 #text change 09-
C/Accession: A91193; A91229; A91387; A91211; A91201; A91200; A43048
R:Rauterberg, J.; Timpl, R.; Furthmayr, H.
Eur. J. Biochem. 27, 231-237, 1972
A/Title: Structural characterization of N-terminal antigenic determi-
A/Reference number: A91193; MUID:72255334; PMID:4115172
A/Accession: A91193
A/Molecule type: protein
A/Residues: 1-19 <RAU>
A/Cross-references: UNIPROT:P02453
A/Experimental source: skin
A/Note: The epsilon carbon of Lys-9, by homology with the rat alpha-
R:Fietzek, P.P.; Kuehn, K.
Eur. J. Biochem. 52, 77-82, 1975
A/Title: The covalent structure of collagen: amino-acid sequence of
A/Reference number: A91229; MUID:76022320; PMID:1164916
A/Accession: A91229
A/Molecule type: protein
A/Residues: 20-145 <FIE>
A/Experimental source: skin
A/Note: Lys-103 is hydroxylated and binds glucosylgalactose
R:Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.
FEBS Lett. 26, 74-76, 1972
A/Title: The covalent structure of collagen: amino acid sequence of
A/Reference number: A91387; MUID:73049499; PMID:4673951
A/Accession: A91387
A/Molecule type: protein
A/Residues: 146-294 <FIZ>
R:Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
Eur. J. Biochem. 38, 396-400, 1973
A/Title: The covalent structure of collagen. 2. The amino-acid sequ-
A/Reference number: A91211; MUID:74086118; PMID:4359390
A/Accession: A91211
A/Molecule type: protein
A/Residues: 295-562 <FIZ>
A/Experimental source: skin
R:Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
Eur. J. Biochem. 30, 169-183, 1972
A/Title: The covalent structure of collagen. The amino-acid sequen-
A/Reference number: A91201; MUID:73042276; PMID:4343808
A/Accession: A91201
A/Molecule type: protein
A/Residues: 563-675 <WEN>
R:Rauterberg, J.; Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
Eur. J. Biochem. 30, 163-168, 1972
A/Title: The covalent structure of collagen. Amino acid sequence of
A/Reference number: A91200; MUID:73042275; PMID:4343807
A/Accession: A91200
A/Molecule type: protein
A/Residues: 676-758 <FIZ>
A/Experimental source: skin
A/Note: Pro-726 is the only 3-hydroxyproline and the only hydroxyla-
R:Rauterberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.;
FEBS Lett. 21, 75-79, 1972
A/Title: The amino acid sequence of the carboxyterminal nonhelical
A/Reference number: A43048
A/Accession: A43048
A/Molecule type: protein
A/Residues: 759-779 <RA2>
A/Experimental source: skin
C/Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658,
C/Comment: Prolines in the third position of the tripeptide repeat

C;Comment: The order of the eight CNBR peptides in the alpha 1(I) chain of bovine skin d
9, 149, 268, and 217 residues.

C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone carboxylic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 54.4%; Score 1898; DB 1; Length 779;
Best Local Similarity 53.0%; Pred. No. 9.6e-89;
Matches 384; Conservative 22; Mismatches 210; Indels 108; Gaps 10;

```

QY 1 GPPGPGPTGLPGPPGGRGPGS-----RGFGADGVAGPKGAGRGSGPGAGKGP 54
Db 29 GPPGARGPQGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 88
QY 55 GEARGPGEAGLPKAGLGTGSPGPPGDKTTPPGAGQDGRPPGPPGARGQAGVMGFP 114
Db 89 GARGLPGTAGLPKMGHGRGSLDAGKADGAPGAPGKPGSPGPNAGAPQMGPRGLPGFP 148
QY 115 GPKGAAGPPKAGRGVPPGPGAVGAPKDGAGAGAGQAGPPGAPGAGRGEGQAGSPGFQ 174
Db 149 GPKGAAGPPKAGRGVPPGPGAVGAPKDGAGAGAGQAGPPGAPGAGRGEGQAGSPGFQ 208
QY 175 GLPGPAGPPGAGKPGQGVFGDLAGPSPGAPG-----PGP-----212
Db 209 GLPGPAGPPGAGKPGQGVFGDLAGPSPGARGRGFPGERGVGPPGAPGRGANGAP 268
QY 213 -----TGLPDPBRRGPGSGRFPFADGVAGPKGAPGERS-----248
Db 269 GNDGAKDAGAPAGSPQAGLQPMGPGERGAAGLPKPKGDRDAGPKADGAPKDGVR 328
QY 249 -----PGPAGKPSGAGRPAGLPGKGLTSGSPGPPGDKTTPPGAGQDGR- 299
Db 329 GLTGPITGPPGAPGDKGAGSPGAGTKGAPDRGEPGPPGAPGAPGADGQPGAK 388
QY 300 -----PGPFPFGARGQAGVMGFPKGAAGEPKKAGRGVPPGAGVGA 345
Db 389 GEPGDAGAKDAGPPGAPGAPGPPGPGIENVGAPKAGSGAGPPGATGFGAAGRVGP 448
QY 346 GKDGAGAGPPGAPGAPG-----RGQAGAPSPGFGPLGPPGAPGAGKPGQGVGDL 402
Db 449 GPGNAGPPGPPGAPGAGKSGRGTGTGAPRGVEGPPGPPGAPGAGADGAPAG 508
QY 403 GAPPGSPGAPGPTGLPDPBRRGPGSGRFPFADGVAGPKGAGRGSGPAGP- - -K 459
Db 509 GTGPPQGIAGQGVVGLPGQGERGFPGLPGSGEPKQKPGSAGSERPPGMPGPPGLA 568
QY 460 GSPGAGRPGEAGLPKAGLGTGSPGPPGDKTTPPGAGQDGRFPGPPGARGQAGVM 519
Db 569 GPPGSGREGAPGAGSGPRDGSQAKGDRGTGAPAGPFPGAPGAPGVGPAKSGDR 628
QY 520 GPPGPKGAAGEPKKAGRGVPPGPGAVGAPKDG-----AGAQPPGP-----563
Db 629 GETGPAGPTGVPVGAARGAPGAPGPPRKBKGTGZGZBGRGKIHGRFGSLQGP 688
QY 564 -----AGPAGRGEGQAGSP---GFGPLGPPGAPGPPGAGKPGEGVPCDLGAPGS 612
Db 689 GEQPSGASGAPGPPGSPAGSPGKGLGLGFLPGIPGPPGRTGADGAPGPPGPP 748
QY 613 GPAG 616
Db 749 GPPG 752

```

RESULT 10

B40333

collagen alpha 1(II) chain precursor - African clawed frog

C;Species: Xenopus laevis (African clawed frog)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: B40333

R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.

J. Cell Biol. 115, 565-575, 1991

A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em

A;Reference number: A40333; MUID:92011898; PMID:1918153

A;Accession: B40333

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1486 <SUA>

A;Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 54.2%; Score 1890.5; DB 1; Length 1486;
Best Local Similarity 54.6%; Pred. No. 3.6e-88;
Matches 374; Conservative 40; Mismatches 196; Indels 75; Gaps 10;

```

QY 1 GPPGPGPTGLPGPPGGRGPGS-----RGFGADGVAGPKGPA-----GERGSGPAGPK 51
Db 269 GPPGPGARGPFGTGLPGVKGHRGYPGLDGAKEAGAGAKGCGATGAGSGPMPGR 328
QY 52 GSPGAGRPGEAGLPKAGLGTGSPGSPGDKTTPPGAGQDGRPPGPPGARGQAGVM 111
Db 329 GLPGERGPPGSSGAAGAR---GNDGLPGP---AGPPGVGAPGAPGPPGSGKEAGPT 382
QY 112 GPPGPKGAAGPPGKAGRGVPPGPGAVGAPKDGAGAGAGQAGPPGAPGAGRGEGQAGSP 171
Db 383 GARGPEGAQGPGRGESGTPGSPGAGASGNGPTDGI PKAGKSSGGPGIAGAPGFPGRGPP 442
QY 172 GFGPLPGPAGPPGAGKPP-----GEOGVPGDLAGAPPS---GPA-----GEPGPT 213
Db 443 GPQATGPLGPKGTGDPFGVAGFKGEGQPKGEISAGPQAGPAGBEGKRGARGEGAA 502
QY 214 GLPFPGERGPGSGRFPFADGVAGPKGAPGERSGPPAGPKGSPGAGRPGEAGLPKAG 273
Db 503 GPNPGERGAPGNRFPQDGLAGPKGAPGERGVPLGPGKGNNGDPGPPGEPGLPEAR 562
QY 274 GLTSGPSGPPDKTTPPGAGQDGRFPGPPGARGQAGVMGFPKGAAGEPKKAGER 333
Db 563 GLTGRPDAGPQGVKPGSGASGEDRPPGPPGARGQAGVGMGFPKGAANGRPKAGEK 622
QY 334 GVPFPKAVGAPGAKDAGAGAGQAGPPGAPGAGRGEGQAGSPGFGPLPGPAGPGEAGKP 393
Db 623 GLVGAPGLGLPKDGTGSGQGNPAGPAGRGEGQAGPPGFPQGLPFGPSPGEGGKP 682
QY 394 GEQVPGDLGAPSPGAPG-----PGPTGLPFPGERGPGSGRFPFADGVAGPKGAPG 450
Db 683 GDOGVPEAGAPGLVGPGRGFRGERSGPPQGLQPRGLPGLTGTGDKGASGSGPN 742
QY 451 GSPGAPGKPSGEGAGRPGEAGLPKAGLGTGSPGSPGDKTTPPGAGQDGRPPGPP 510
Db 743 GAQFPGLQGMPPGERGAAGISGPKGDRGDTGKGPESGASGKDSRGLTGTGIPPGAPGN 802
QY 511 GARGQAGVMGFPKGAAGEPKKAGRGVPPGPGAVGAPGAKD-----GEA 555
Db 803 GEKESGSPSPGPIVGARGAPGDRGENGPPGAPGAPGADQSGGLKDGQSGSGKGD 862
QY 556 GAQPPGPA-----GPAGRGEGQAGSPGFGQL-----PGPAGPP 591
Db 863 GAFPGQPSGAPGPPGPTGTVFGPKGARGAQQAGATGFGCAAGRVTGTPGNGNPPGPP 922
QY 592 GEAGKPEQGVPCDLGAPGSPGAP 616
Db 923 GSAGKEGPKVRGDAGPPGRAGDPG 947

```

RESULT 11

A40333

collagen alpha 1'(II) chain precursor - African clawed frog

C;Species: Xenopus laevis (African clawed frog)

C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004

C;Accession: A40333

R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.

J. Cell Biol. 115, 565-575, 1991

A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em

QY 1 GPFGPGPTGLPGPPGERGCGPSRGFPFGADGVAGPKPGAGRGSPGAPGPGSGAGRP 60
Db 468 GSPGPGAGNLPGAAGGAPGFGPAGPAGNIGPKPGAGRGAPGAPGAPGAPGAGPGRD 527

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:17:54 ; Search time 91.5221 Seconds
(without alignment)
3452.202 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGEPGTGLPGPPGRRGG.....GEQGVPGDLGAPGSPGAGG 617

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	2237.5	64.1	1 1464 1 CALL_HUMAN	P02452 homo sapien
2	2233.5	64.0	2 1461 2 O76045	O76045 homo sapien
3	2233.5	64.0	1464 2 Q8N473	Q8N473 homo sapien
4	2231.5	64.0	1460 1 CALL_CANFA	Q9X8J7 canis famil
5	2222	63.7	1069 2 Q6LAN8	Q6LAN8 homo sapien
6	2215.5	63.5	1453 2 Q63079	Q63079 rattus norv
7	2205.5	63.2	1453 1 CALL_MOUSE	P11087 mus musculu
8	2205.5	63.2	1453 2 Q810J9	P02452 mus musculu
9	2176.5	62.4	1225 2 Q6PCL3	Q6PCL3 mus musculu
10	2151.5	61.7	1453 1 CALL_CHICK	P02457 gallus gall
11	2043.5	58.6	1450 2 O9VIB4	O9VIB4 cynops pyrr
12	2042.5	58.6	1445 2 O93251	O93251 rana catesb
13	2037.5	58.4	1449 2 Q640B2	Q640B2 xenopus tro
14	2029	58.2	1449 2 Q802B5	Q802B5 xenopus lae
15	2010.5	57.6	671 1 CALL_RAT	P01B91 rattus norv
16	1976.5	56.7	1447 2 O9IB91	O9IB91 xenopus lae
17	1953	56.0	1160 2 Q14046	Q14046 homo sapien
18	1953	56.0	1487 2 Q14047	Q14047 homo sapien
19	1952	56.0	1487 2 Q77753	Q77753 canis famil
20	1944	55.7	1418 2 Q828396	Q828396 equus cabal
21	1943	55.7	1418 1 CALL2_HUMAN	P02458 homo sapien
22	1942	55.7	1420 2 Q90W37	Q90W37 gallus gall
23	1937	55.5	1269 2 Q77227	Q77227 gallus gall
24	1936	55.5	1442 2 Q62031	Q62031 mus musculu
25	1936	55.5	1442 2 Q62033	Q62033 mus musculu
26	1936	55.5	1459 1 CALL2_MOUSE	P28481 mus musculu
27	1936	55.5	1459 2 Q62032	Q62032 mus musculu
28	1927	55.2	1419 2 Q80VY3	Q80VY3 mus musculu
29	1927	55.2	1419 2 Q80X38	Q80X38 mus musculu
30	1927	55.2	1487 2 Q641K3	Q641K3 mus musculu
31	1923	55.1	1447 2 Q6U1J5	Q6U1J5 brachydanio

32	1921	55.1	1419	2	Q63123	Q63123 rattus norv
33	1914	54.9	1447	2	O6P4U1	O6P4U1 brachydanio
34	1898	54.4	779	1	CALL_BOVIN	P02453 bos taurus
35	1892	54.2	1492	2	Q6P4Z2	Q6P4Z2 xenopus tro
36	1890.5	54.2	1486	2	Q91717	Q91717 xenopus lae
37	1888.5	54.1	1486	2	Q7ZTI6	Q7ZTI6 xenopus lae
38	1883	54.0	1418	2	Q9W7R9	Q9W7R9 cynops pyrr
39	1880	53.9	1491	2	Q91718	Q91718 xenopus lae
40	1880	53.9	1491	2	Q7ZTM4	Q7ZTM4 xenopus lae
41	1873	53.7	1262	1	CALL_CHICK	P12105 gallus gall
42	1873	53.7	1449	2	Q910C0	Q910C0 oncothynchu
43	1860	53.3	1449	2	Q6PEI9	Q6PEI9 brachydanio
44	1858	53.3	1449	2	Q6NZ15	Q6NZ15 brachydanio
45	1856	53.2	1049	1	CALL_BOVIN	P04258 bos taurus

ALIGNMENTS

RESULT 1
CALL_HUMAN
ID CALL_HUMAN STANDARD; PRT; 1464 AA.
AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
AC Q15201; Q16050; Q7KZ30; Q7KZ34; Q8IVI5; Q9UML6; Q9UWM7;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT SER-1434.
RA Dalgleish R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-589 FROM N.A.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RX D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RX Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RX Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RX Jaenisch R., Prockup D.J.;
RA "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RX Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RX Myers J., Williams C., Ramirez F.;
RA "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [5]
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RX Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [6]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;

RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry B.F.,
RA Prockop D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT of alpha 1 chain of human type I procollagen. Statistical evaluation
RT of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN [7]
RP SEQUENCE OF 472-607 FROM N.A.
RX PubMed=2981843;
RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
RT "Multixen deletion in an osteogenesis imperfecta variant with
RT increased type III collagen mRNA.";
RL J. Biol. Chem. 260:691-694(1985).
RN [8]
RP SEQUENCE OF 488-625 FROM N.A.
RX PubMed=3857621;
RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
RT I collagen chain in a lethal form of osteogenesis imperfecta.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
RN [9]
RP SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OF-II HIS-1277; ARG-1388 AND
RP 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
RX MEDLINE=93352646; PubMed=8349697;
RA Chessler S.D., Wallis G.A., Byers P.H.;
RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
RT chain of type I collagen result in defective chain association and
RT produce lethal osteogenesis imperfecta.";
RL J. Biol. Chem. 268:18218-18225(1993).
RN [10]
RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=88124208; PubMed=3340531;
RA Maekelae J.K., Raessina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [11]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [12]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to transcriptional
RT control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [13]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [14]
RP SEQUENCE OF 33-52.
RX PubMed=2318855;
RA Wirtz M.K., Keene D.R., Hori H., Glanville R.W., Steinmann B.,
RA Rao V.H., Hollister D.W.;
RT "In vivo and in vitro noncovalent association of excised alpha 1 (I)
RT amino-terminal propeptides with mutant alpha 2(I) collagen chains
RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
RT VII.";
RL J. Biol. Chem. 265:6312-6317(1990).
RN [15]
RP SEQUENCE OF 156-183 FROM N.A.
RX PubMed=2767050;
RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,

RA Bateman J.F.;
RT "A base substitution in the exon of a collagen gene causes alternative
RT splicing and generates a structurally abnormal polypeptide in a
RT patient with Ehlers-Danlos syndrome type VII.";
RL EMBO J. 8:1705-1710(1989).
RN [16]
RP SEQUENCE OF 175-187 AND 274-289.
RX PubMed=2169412;
RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
RA Muller P.K.;
RT "A critical crosslink region in human-bone-derived collagen type I.
RT Specific cleavage site at residue Leu95.";
RL Eur. J. Biochem. 192:153-159(1990).
RN [17]
RP SEQUENCE OF 263-268.
RC TISSUE=Skin;
RX MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT "A comparative study of glycopeptides derived from selected vertebrate
RT collagens. A possible role of the carbohydrate in fibril formation.";
RL J. Biol. Chem. 245:5042-5048(1970).
RN [18]
RP SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND
RP 1143-1162 FROM N.A.
RX PubMed=2374517;
RA Labhard M.E., Hollister D.W.;
RT "Segmental amplification of the entire helical and telopeptide regions
RT of the cDNA for human alpha 1 (I) collagen.";
RL Matrix 10:124-130(1990).
RN [19]
RP SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
RX MEDLINE=80364528; PubMed=6183642;
RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
RT "Cloning and characterization of five overlapping cDNAs specific for
RT the human pro alpha 1(I) collagen chain.";
RL Nucleic Acids Res. 10:5925-5934(1982).
RN [20]
RP SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
RX PubMed=2339700;
RA Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
RT "Variable expression of osteogenesis imperfecta in a nuclear family is
RT explained by somatic mosaicism for a lethal point mutation in the
RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
RL Am. J. Hum. Genet. 46:1034-1040(1990).
RN [21]
RP SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
RX MEDLINE=95187161; PubMed=7881420;
RA Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
RA Brunelli P.C., Mottes M.;
RT "Severe (type III) osteogenesis imperfecta due to glycine
RT substitutions in the central domain of the collagen triple helix.";
RL Hum. Mol. Genet. 3:2201-2206(1994).
RN [22]
RP SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.
RX PubMed=3170557;
RA Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andreassen P.,
RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
RT "Substitution of cysteine for glycine within the carboxyl-terminal
RT telopeptide of the alpha 1 chain of type I collagen produces mild
RT osteogenesis imperfecta.";
RL J. Biol. Chem. 263:14605-14607(1988).
RN [23]
RP SEQUENCE OF 1440-1464 FROM N.A.
RX MEDLINE=90110490; PubMed=2295701;
RA Willing M.C., Cohn D.H., Byers P.H.;
RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I
RT collagen predicts an elongated pro alpha 1(I) chain and results in
RT osteogenesis imperfecta type I.";
RL J. Clin. Invest. 85:282-290(1990).
RN [24]
RP SEQUENCE OF 1454-1464 FROM N.A.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;

RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";

Query Match 64.1%; Score 2237.5; DB 1; Length 1464;
Best Local Similarity 60.2%; Pred. No. 3.6e-81;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEGPTGLCPGCGGCGSGRGGFPGADGVAGPKGAGRGSGCPGA 48
DB 458 GPPGAGBEGKRGARGEGPTGLPFGGCGGSGRGGFPGADGVAGPKGAGRGSGCPGA 517
QY 49 GPKGSPGAGRGAGLPGAKGLTSGSGPGDGKTPGPGAGQDGRGPPGPPGARGQA 108
DB 518 GPKGSPGAGRGAGLPGAKGLTSGSGPGDGKTPGPGAGQDGRGPPGPPGARGQA 577
QY 109 GVMGPPGKGAAGEGKAGRGVPPGCAVGPAGKDGAGAGQGGPPGAPGAGEGQCPA 168
DB 578 GVMGPPGKGAAGEGKAGRGVPPGCAVGPAGKDGAGAGQGGPPGAPGAGEGQCPA 637
QY 169 GSPGQGLPGAPGPGGAGKPGGQGVPGDLGAPGSGPAGE-----GPG----- 212
DB 638 GSPGQGLPGAPGPGGAGKPGGQGVPGDLGAPGSGPAGE-----GPG----- 697
QY 213 -----TGLPGPCGCGSGRGGFPGADGVAGPKGAGRGSGCPGAPGKSP 258
DB 698 GAGAPGNDGAKGADGAPGAGCGSGLQGMPGGERGAAGLPKGRDGRDAGPKGADGSP 757
QY 259 GE-----AGRPGGAGLPGAK-----GLTSGSGSGPPGDKTTPGPGAGQDGRP 300
DB 758 GKGVRGLTGTGTPGPGAGADKSGSGSPGAGTARGAPGDRGEPGPGPAGFAGPP 817
QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGEGKAGRGVPPGPPGA----- 341
DB 818 GADGQFGAKGEGPDAGAKGDAGPPGAPGAPGPGPIGNVGAAPGAKGARGSGAPGATGFP 877
QY 342 -----VGPAGKDGAGAGGPPGAPG-----AGRGQGGAGSGPFGGLGPPGAPGCEAGKP 393
DB 878 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGETGAPRGVGGPPGPPGAPGKSP 937
QY 394 GPGQVPGDLGPPGSPGAGEPGTGLPAPPGRGGGSGRGGFPGADGVAGPKGAGRGSP 453
DB 938 GADGAPAGPTGPGQIAGQGVGLPQGRGEGFGLPGSGEGPKQSGSGASGERGPP 997
QY 454 GPAGP---KGSFGGAGRGAGLPGAKGLTSGSGSPGDKTTPGPGAGQDGRGPPGPP 510
DB 998 GPMGPPGLAGPPGSGREGAPGAGSGRSGDPSGAKGDRGETGAPGPPGAPGAPGAPV 1057
QY 511 GARGQAGVMGPPGKGAAGEGKAGRGVPPGPPGAGVAGPKDGE-----AGAQ 558
DB 1058 GPAGKSGDRGETGAPGAPGVPVGPAGPAGPAGPQGGPRGDKGETGQGGDRGKGHRGFGSLQ 1117
QY 559 GPPGP-----AGPAGRGEGQGPAGSP---GFGGLPGAPGPPGAGKPGCEQGV 603
DB 1118 GPPGPPSGEGQGPAGSPGAPGPRGPPGAGAPGKDGGLNGLPGLPGLPGRGRTGDGAPV 1177
QY 604 GDLGAPGSPGAPG 616
DB 1178 GPPGPPGPPGPPG 1190

RESULT 2

ID O76045 PRELIMINARY; PRT; 1461 AA.
AC DT O76045;
CT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A.
RP Chu M.L., de Wet W., Bernard M., Ramirez F.;
RA MEDLINE=85130970; PubMed=2857713;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
[2] SEQUENCE FROM N.A.
RP MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
[3] SEQUENCE FROM N.A.
RP MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
[4] SEQUENCE FROM N.A.
RP MEDLINE=9113770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
[5] SEQUENCE FROM N.A.
RP MEDLINE=92157916; PubMed=1787829;
RA Westershausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
[6] SEQUENCE FROM N.A.
RP MEDLINE=98107942; PubMed=9443882;
RA Korhko J., Ala-Kokko L., De Paepe A., Nuytinck L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:
RT identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
[7] SEQUENCE FROM N.A.
RP Korhko J.M., Earley J.J., Nuytinck L., DePaepe A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR GO; GO:0005581; C:cytoplasm; IEA.
DR GO; GO:0005737; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SCCI.
DR InterPro; IPR01007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWFC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS50184; VWFC_2; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;


```

Query Match      64.0%; Score 2233.5; DB 2; Length 1461;
Best Local Similarity 60.0%; Pred. No. 5.2e-81;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGRGSGRGGPGADGVAGPKGAGERSGPPA 48
D 455 GPPGAGEGKRGARGEPGPTGLPGRGSGRGGPGADGVAGPKGAGERSGPPA 514
QY 49 GPKSGPAGRPGAGLPGAKGLTSGPSGPGDKTTPPGAGQDGRPPGPPGARGQA 108
D 515 GPKSGPAGRPGAGLPGAKGLTSGPSGPGDKTTPPGAGQDGRPPGPPGARGQA 574
QY 109 GVMGFPKGAAGPCKAGRGVPGPCAVGAPGAKGAGAGQPPGAGRGSGGPA 168
D 575 GVMGFPKGAAGPCKAGRGVPGPCAVGAPGAKGAGAGQPPGAGRGSGGPA 634
QY 169 GSPFGQLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
D 635 GSPFGQLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 694
QY 213 -----TGLPGRGSGRGGPGADGVAGPKGAGERSGPPGAGKGP 258
D 695 GAGAPGNDGAKGADGAGAPGAGSGAPGQAGPQAGPAGTGAAGPDRGEPGPPGAGPP 754
QY 259 GE-----AGRPGAGLPGAK-----GLTSGPSGPGDKTTPPGAGQDGRP 300
D 755 GKDGVRGLTGPPIPGPAGAPGDKGESGSPGAGTGAAGPDRGEPGPPGAGPP 814
QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGPCKAGRGVPGPPGA----- 341
D 815 GADQPKAGPAGDAGAKGADGPPGAPGAPGPGTGNVGAAGKAGSAGPGATGPP 874
QY 342 -----VGPAGKDGAGAGQPPGAGP---AGRGSGPAGSGPFGQLPGPAGPPGAGKP 393
D 875 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGPGTGPAGPGEVPPGPPGAGKGP 934
QY 394 GEQGVPGDLGAPGSGPAGPPTGLPGRGSGRGGPGADGVAGPKGAGERSG 453
D 935 GADGAPAGTPTGPGIAGRGVVGLPQGRGGRGFFGLPGRGSGRGGPGAG 994
QY 454 GPAGP---KSGPAGRPGAGLPGAKGLTSGPSGPGDKTTPPGAGQDGRPPGPP 510
D 995 GPMGPPGLAGPAGESREGAPGAGSGRGGPGAGKDRGEGTGPAGPPGAPGAPGV 1054
QY 511 GARGOAGVMGFPKGAAGPCKAGRGVPGPCAVGAPGAGKGE-----AGAQ 558
D 1055 GPAGKSGDRGTGPAGPAGPVGARGPAGPQGRGKGTGEGQDGRGKGRGFGSLQ 1114
QY 559 GPPGP-----AGPAGRGGQPGAGSP---GFGQLPGPAGPPGAGKPGEGQVP 603
D 1115 GPPGPPGSGEGQPGSAGSGPAGPAGPPGAGKDRGKGLNGLPQIPGPPGRGTGAGPV 1174
QY 604 GD/LGAPGSPGAG 616
D 1175 GPPGPPGPPGPPG 1187

RESULT 3
Q8N473
ID Q8N473 PRELIMINARY; PRT; 1464 AA.
AC Q8N473
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen, preproprotein.
GN Name=COL1A1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;

```

```

MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
SRAusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
Blakesley R.W., Young A.C., Shevchenko Y., Bouffard G.G.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszinski M.L., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA SRAusberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SCCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clq helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS01084; VWF_C_2; 1.
KW Collagen.
SQ
SEQUENCE 1464 AA; 139010 MW; B0581F8D1C89DDE8 CRC64;

Query Match      64.0%; Score 2233.5; DB 2; Length 1464;
Best Local Similarity 60.0%; Pred. No. 5.2e-81;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGRGSGRGGPGADGVAGPKGAGERSGPPA 48
D 458 GPPGAGEGKRGARGEPGPTGLPGRGSGRGGPGADGVAGPKGAGERSGPPA 517
QY 49 GPKSGPAGRPGAGLPGAKGLTSGPSGPGDKTTPPGAGQDGRPPGPPGARGQA 108
D 518 GPKSGPAGRPGAGLPGAKGLTSGPSGPGDKTTPPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGFPKGAAGPCKAGRGVPGPCAVGAPGAKGAGAGQPPGAGRGSGGPA 168
D 578 GVMGFPKGAAGPCKAGRGVPGPCAVGAPGAKGAGAGQPPGAGRGSGGPA 637
QY 169 GSPFGQLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
D 638 GSPFGQLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 697
QY 213 -----TGLPGRGSGRGGPGADGVAGPKGAGERSGPPGAGKGP 258
D 698 GAGAPGNDGAKGADGAGAPGAGSGAPGQAGPQAGPAGTGAAGPDRGEPGPPGAGPP 757
QY 259 GE-----AGRPGAGLPGAK-----GLTSGPSGPGDKTTPPGAGQDGRP 300
D 759 GAGAPGNDGAKGADGAGAPGAGSGAPGQAGPQAGPAGTGAAGPDRGEPGPPGAGPP 814

```



```

Db 758 GKGVRLTGLTPIPPGAPAGDKGESGSPGAGTGTGARGAPGDRGPPGPPGAGPAP 817
Qy 301 GPPGPPGARGQ-----AGVMGFPKGAGBPAGKAGRGVPPGPA----- 341
Db 818 GADQCPKAGBPAGAGKAGDAGPPGAPGPPGIGNVGAPGAKGARGSGAPGATGFP 877
Qy 342 -----VGPAGKDGAGAGGPPGAPG---ACERGEQGPAGSPGFGQLPGPAGPPGAGK 393
Db 878 GAARVVPSPGSGNAGPPGPPGAPGKGGKGRGTGAPGPPGPPGPPGAPGK 937
Qy 394 GEOQVPGDLGAPSGPAGPPTGLPFPGERGGPGRGPPGADGAGVAGPKGPPGAGRGSP 453
Db 938 GADGAPAGPTGPGQIAGRGVGLPGQRCGERGFPGLPGSPGPKGPGSGASGERGPP 997
Qy 454 GPAGP-----KGSFGAGRGAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRGPPGPP 510
Db 998 GPMGPPGLAGPPGSGREGAPGABGSPGRDGSAPGAKGDRGTGPPGPPGAPGAPG 1057
Qy 511 GARGOAGVMGPPGPKGAGBPAGKAGRGVPPGAPGAGKDG-----AGAQ 558
Db 1058 GPAGSGDRGTGAPGAPGPPGAPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPP 1117
Qy 559 GPPGP-----AGPAGRGEGQGPAGSP---GFQGLPGPAGPPGAGKPGEGQGP 603
Db 1118 GPPGPPGSGQSGASGAPGPPGASAPGKGLGLGPPGPPGPPGPPGPPGPPGPP 1177
Qy 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 4
CALL CANFA
ID CALL CANFA STANDARD; PRT; 1460 AA.
AC Q9XSJ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 25-OCT-2004 (Rel. 45, Last annotation update)
RT "Sequence of normal canine COL1A1 cDNA and identification of a heterozygous alpha 1(I) collagen Gly208Ala mutation in a severe case of canine osteogenesis imperfecta."
GN Name=COL1A1;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
RC TISSUE=Skin;
RX MEDLINE=2102337; PubMed=11147834; DOI=10.1006/abbi.2000.2099;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA and identification of a heterozygous alpha 1(I) collagen Gly208Ala mutation in a severe case of canine osteogenesis imperfecta."
RL Arch. Biochem. Biophys. 384:37-46 (2000).
CC -1- FUNCTION: Type I collagen is a member of group I collagen (fibrillar forming collagen).
CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -1- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta (OI).
CC -1- SIMILARITY: Belongs to the fibrillar collagen family.
CC -1- SIMILARITY: Contains 1 WFCC domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF153062; AAD34619.1; -.
DR InterPro; IPR008161; Cig_helix.

```

```

DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008885; Fib_collagen_C.
DR InterPro; IPR001007; WMF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Cig_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR PROSITE; PS01208; WMF_C_1; 1.
DR PROSITE; PS01208; WMF_C_2; 1.
DR PROSITE; PS01208; WMF_C_3; 1.
KW Collagen; Disease mutation; Extracellular matrix; Glycoprotein;
KW Hydroxylation; Pyrrolidone carboxylic acid; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 22 By similarity.
FT PROPEP 23 157 N-terminal propeptide.
FT CHAIN 158 1214 Collagen alpha 1(I) chain.
FT PROPEP 1215 1460 C-terminal propeptide.
FT DOMAIN 34 92 WMF.
FT DOMAIN 158 174 Nonhelical region (N-terminal).
FT DOMAIN 175 1188 Triple-helical region.
FT DOMAIN 1189 1214 Nonhelical region (C-terminal).
FT SITE 741 743 Cell attachment site (Potential).
FT SITE 1089 1091 Cell attachment site (Potential).
FT MOD_RES 158 158 Pyrrolidone carboxylic acid (By similarity).
FT MOD_RES 166 166 Allysine (By similarity).
FT MOD_RES 261 261 5-hydroxylysine (By similarity).
FT MOD_RES 1160 1160 3-hydroxyproline (By similarity).
FT CARBOHYD 261 261 O-linked (Gal.. ) (By similarity).
FT CARBOHYD 1361 1361 N-linked (GlcNAc.. ) (By similarity).
FT VARIANT 208 208 G -> A (in OI; severe).
FT SEQUENCE 1460 AA; 138762 MW; 58E367AD2B570697 CRC64;

Query Match 64.0%; Score 2231.5; DB 1; Length 1460;
Best Local Similarity 60.2%; Pred. No. 6.2e-81;
Matches 441; Conservative 20; Mismatches 155; Indels 117; Gaps 12;

Qy 1 GPP-----GEPGPTGLPGERGPGSGRFPGADGAGVAGPKGAGERGSGCPA 48
Db 454 GPPGAPGEGKRGARGPPTGLPGERGPGSGRFPGADGAGVAGPKGAGERGSGCPA 513
Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRGPPGPPGARGQA 108
Db 514 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRGPPGPPGARGQA 573
Qy 109 GVMGFPKGAGBPAGKAGRGVPPGAPGAPGAGKDGAGAGQGPAGPAGRGEGQCPA 168
Db 574 GVMGFPKGAGBPAGKAGRGVPPGAPGAPGAGKDGAGAGQGPAGPAGRGEGQCPA 633
Qy 169 GSPGFQGLPGPAGPPGAGKRGVPPGAGVPPGAGVPPGAGVPPGAGVPPGAGVPP 212
Db 634 GSPGFQGLPGPAGPPGAGKRGVPPGAGVPPGAGVPPGAGVPPGAGVPPGAGVPP 693
Qy 213 -----TGLPGERGPGSGRFPGADGAGVAGPKGAGERGSGCPAGPKGSP 258
Db 694 GAGAPNDGAKGADGAPGAPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 753
Qy 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGPDGKTGPPGAGQDGRP 300
Db 754 GKGVRLTGLTPIPPGAPAGDKGESGSPGAGTGTGARGAPGDRGPPGPPGAGPAP 813
Qy 301 GPPGPPGARGQ-----AGVMGFPKGAGBPAGKAGRGVPPGPA----- 341
Db 814 GADQCPKAGBPAGAGKAGDAGPPGAPGPPGIGNVGAPGAKGARGSGAPGATGFP 873
Qy 342 -----VGPAGKDGAGAGGPPGAPG---ACERGEQGPAGSPGFGQLPGPAGPPGAGK 393
Db 874 GAARVVPSPGSGNAGPPGPPGAPGKGGKGRGTGAPGPPGPPGPPGAPGK 933
Qy 394 GEOQVPGDLGAPSGPAGPPTGLPFPGERGGPGRGPPGADGAGVAGPKGPPGAGRGSP 453
Db 938 GADGAPAGPTGPGQIAGRGVGLPGQRCGERGFPGLPGSPGPKGPGSGASGERGPP 993
Qy 454 GPAGP-----KGSFGAGRGAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRGPPGPP 510

```


RESULT 7

CALL MOUSE

ID CALL MOUSE STANDARD; PRT; 1453 AA.

AC P11087; Q60635;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 25-OCT-2004 (Rel. 45, Last annotation update)

DE Collagen alpha 1(I) chain precursor.

DE Names=Colla1; Synonyms=Colal;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

[1]

RP SEQUENCE FROM N.A.

RP STRAIN=FVB/N;

RC MEDLINE=96033240; PubMed=8535610;

RA Li S.W., Killian J., Prockop D.J.;

RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain

RT of type I procollagen.";

RL Matrix Biol. 14:593-595(1995).

[2]

RP SEQUENCE OF 518-1128 FROM N.A.

RA MEDLINE=8637403; PubMed=3841523; DOI=10.1016/0378-1119(85)90329-4;

RC French B.T., Lee W.-H., Maul G.G.;

RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I) collagen

RT protein.";

RL Gene 39:311-312(1985).

[3]

RP SEQUENCE OF 735-1130 FROM N.A.

RA MEDLINE=83141374; PubMed=6298597;

RC Monson J.M., Friedman J., McCarthy B.J.;

RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:

RT evidence for a mouse B1 element within the gene.";

RL Mol. Cell. Biol. 2:1362-1371(1982).

[4]

RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.

RA MEDLINE=83157109; PubMed=6219867;

RC Monson J.M., McCarthy B.J.;

RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:

RT evidence for insertions or deletions in gene coding sequences.";

RL DNA 1:59-69(1981).

[5]

RP SEQUENCE OF 1442-1453 FROM N.A.

RA MEDLINE=88124276; PubMed=3340560;

RC Mooslehner K., Harbers K.;

RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size of

RT the 3'-untranslated region.";

RL Nucleic Acids Res. 16:773-773(1988).

CC -1- FUNCTION: Type I collagen is a member of group I collagen

CC (fibrillar forming collagen).

CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.

CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and

CC bones. In bones the fibrils are mineralized with calcium

CC hydroxyapatite.

CC -1- PTM: Prolines at the third position of the tripeptide repeating

CC unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -1- SIMILARITY: Belongs to the fibrillar collagen family.

CC -1- SIMILARITY: Contains 1 WFPC domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on

CC use by non-profit institutions as long as its content is in no

CC modified and this statement is not removed. Usage by and for commer

CC entities requires a license agreement (See <http://www.isb-sib.ch/announc>

CC or send an email to license@isb-sib.ch).

CC

DR ENBL; U08020; AAA88912.1; -

DR ENBL; X15896; CAA33904.1; -

DR ENBL; M14423; AAA37333.1; -

DR ENBL; M17491; AAA37334.1; -

DR ENBL; X06753; CAA29927.1; -

DR	EMBL; K03036; AAA37332.1; ..	
DR	EMBL; K03029; AAA37332.1; JOINED.	
DR	EMBL; K03030; AAA37332.1; JOINED.	
DR	EMBL; K03031; AAA37332.1; JOINED.	
DR	EMBL; K03032; AAA37332.1; JOINED.	
DR	EMBL; K03033; AAA37332.1; JOINED.	
DR	EMBL; K03034; AAA37332.1; JOINED.	
DR	EMBL; K03035; AAA37332.1; JOINED.	
DR	PIR; S57243; S21626.	
DR	MGI; 88467; Col1a1.	
DR	InterPro; IPR008161; Clg helix.	
DR	InterPro; IPR008180; Collagen.	
DR	InterPro; IPR00885; Fib collagen_C.	
DR	InterPro; IPR009041; PMP SGC1.	
DR	InterPro; IPR001007; VWF C.	
DR	Pfam; PF01410; COLFI; 1.	
DR	Pfam; PF01391; Collagen; 18.	
DR	ProDom; PD000007; Clg_helix; 3.	
DR	SMART; SM002078; Fib collagen_C; 1.	
DR	SMART; SM00038; COLFI; 1.	
DR	SMART; SM00214; VWC; 1.	
DR	PROSITE; PS01208; VWF C; 1.	
DR	PROSITE; PS0184; VWF C; 2.	
KW	Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;	
KW	Pyridinolone carboxylic acid; Repeat; Signal; Structural protein.	
FT	SIGNAL 1 22	
FT	PROPEP 23 151	N-terminal propeptide.
FT	CHAIN 152 1207	Collagen alpha 1(I) chain.
FT	PROPEP 1208 1453	C-terminal propeptide.
FT	DOMAIN 29 87	VWFC
FT	DOMAIN 152 167	Nonhelical region (N-terminal).
FT	DOMAIN 168 1181	Triple-helical region.
FT	DOMAIN 1182 1207	Nonhelical region (C-terminal).
FT	MOD_RES 152 152	Pyridinolone carboxylic acid (By similarity).
FT	MOD_RES 160 160	Allysine (By similarity).
FT	MOD_RES 254 234	5-hydroxylysine (By similarity).
FT	MOD_RES 1153 1153	3-hydroxyproline (By similarity).
FT	CARBOHYD 56 56	N-linked (GlcNAc..) (Potential).
FT	CARBOHYD 254 254	O-linked (Gal..) (By similarity).
FT	CARBOHYD 1354 1354	N-linked (GlcNAc..) (By similarity).
FT	SITE 734 736	Cell attachment site (Potential).
FT	SITE 1082 1084	Cell attachment site (Potential).
FT	CONFLICT 1450 1450	A -> V (in Ref. 5).
FT	SEQUENCE 1453 AA; 137944 MW; 3B802E535DF81808 CRC64;	
Query Match	63.2%; Score 2205.5; DB 1; Length 1453;	
Best Local Similarity	58.9%; Pred. No. 6.6e-80;	
Matches 432; Conservative 22; Mismatches 162; Indels 117; Gaps 11;		
QY	1 GPP-----GEPQTLPGPPGGRGGRGPPGADGVAGPKGAGERGSPGA 48	
Db	447 GPPGAGEGKRGAGEGPPGGLPGPPGGRGGRGPPGADGVAGPKGAGERGSPGA 506	
QY	49 GPKSGPGEAGRGEAGLPCAKGLTSPGSPGDKTTPPGAGQGRGPPGPGARQQA 108	
Db	507 GPKSGPGEAGRGEAGLPCAKGLTSPGSPGDKTTPPGAGQGRGPPGPGARQQA 566	
QY	109 GVMGPPGPKGAAGEGPKGAGERGVPGPAGVAGKDGAGAGQPPGAPGAGEGQGA 168	
Db	567 GVMGPPGPKGAGEGPKGAGERGVPGPAGVAGKDGAGAGQPPGAPGAGEGQGA 626	
QY	169 GSPFGQLPGPAPGPEAGKPGEQGVGDLGAPGSPGAGE-----PGP---- 212	
Db	627 GSEGFQGLPGPAPGPEAGKPGEQGVGDLGAPGSPGAGEFPGERGVPQPPGAPGR 686	
QY	213 -----TCLPPGPPGGRGGRGPPGADGVAGPKGAGERGSPGAPGKSP 258	
Db	687 GNNGAPGNDGAKGDTGAPGAPGSGQAGLQGMFGERGAAGLPGPKGDRGDAGPKGADGSP 746	
QY	259 GE-----AGRPGEAGLPCAKGLTSPGSPGPD-----GKTGPPGPGAGDGRP 300	
Db	747 GKDAGRLTGPIGPPGAPAGDPKGAEGPSGPPGTGARGAPGDRGAGPPGAPGAPFP 806	

QY	301 GPPPPGARGQ-----AGVMGPPGPKGAAGBPBKAGRGVPPGPPGA----- 341	
Db	807 GADGQPGAKGPPGDTGVKGDAGPPGAPGAPGPPGIGNVGAPGPKPGAGAPPGATGPP 866	
QY	342 -----VGPAGKDGAGAGQPPGAPG-----AGERGQGPAGSPGFGGLPGPAGPGEAGKP 393	
Db	867 GAAGRVGPPGSGNAGPPGPPGPGVKGEGKGRGTGAPRGPGEVPPGPPGAPGAGEKSP 926	
QY	394 GEQGVGDLGAPGSPGAPGEPGTLPGPPGGRGGRGPPGADGVAGPKGAGERGSP 453	
Db	927 GADGAPGSPGTPGQGTAGVGVGLPGQGRGPPGPPGPPGSGEPGKQSPGSSGERGPP 986	
QY	454 GPAGP---KGSPPGEAGRGEAGLPCAKGLTSPGSPGDKTTPPGAGQGRGPPGPP 510	
Db	987 GPMGPPGLAGPPGSGREGSPGAEGRGPDGAPGAKGDRGTGTPAGPPGAPGAPGAPGV 1046	
QY	511 GARGQAGVMGPPGPKGAAGEGPKGAGERGVPGP-----PGAV 546	
Db	1047 GPAGKNGDRGTGTPAGPAGPITGPAGAPGPPGPPGDKGTGEQCDRGKIHGRFGSLQ 1106	
QY	547 GPAGKDGAGAGQPPGPPGAPGAGEGQGPAGSP---GQGLPGPAGPGEAGKPGEGQVP 603	
Db	1107 GPPSGPSGEGSGASGAPGPPGPPGSGSPGKDLGLGLPGIPGPPGPPGRTGDSGA 1166	
QY	604 GDLGAPGSPGAPG 616	
Db	1167 GPPGPPGPPGPPG 1179	
RESULT 8		
Q810J9	PRELIMINARY; PRT; 1453 AA.	
ID	Q810J9	
AC	Q810J9	
DT	01-JUN-2003 (TrEMBLrel. 24, Created)	
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Procollagen, type I, alpha 1.	
GN	Name=Col1a1;	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RC	STRAIN=FVB/N; TISSUE=Colon;	
EX	MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;	
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,	
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,	
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,	
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,	
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,	
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,	
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,	
RA	Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,	
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,	
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,	
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,	
RA	Whiting R., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,	
RA	Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,	
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,	
RA	Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E.,	
RA	Jones S.J., Marra M.A.	
RT	"Generation and initial analysis of more than 15,000 full-length human	
RT	and mouse cDNA sequences."	
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FVB/N; TISSUE=Colon;	
RA	Klausner R.;	
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; BC050014; AA050014.1; -.	

Db 1167 GPPGPPGPPGPPG 1179

RESULT 9

Q6PCL3 PRELIMINARY; PRT; 1225 AA.

AC Q6PCL3;

AC Q6PCL3;

DT 05-JUL-2004 (TEMBLrel. 27, Created)

DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)

DE Collal protein.

GN Names:Collal;

OS Mus musculus (Mouse).

OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ON NCBI_TaxID=10090;

RX [1]

RC SEQUENCE FROM N.A.

RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,

RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP [2]

RC SEQUENCE FROM N.A.

RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;

RA Strausberg R.;

RA Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.

RL EMBL: BC059281; AAH59281.1; -

DR GO: GO:0005615; C:extracellular space; TAS.

DR InterPro: IPR008161; Clg_helix.

DR InterPro: IPR008160; Collagen.

DR InterPro: IPR000885; Fib_collagen_C.

DR InterPro: IPR009041; PMP_SGGCI.

DR InterPro: IPR001007; VWF_C.

DR Pfam: PF01410; COLFI; 1.

DR Pfam: PF01391; Collagen; 13.

DR Pfam: PF00093; VWC; 1.

DR ProDom: PD000007; Clg_helix; 2.

DR ProDom: PD002078; Fib_collagen_C; 1.

DR SMART: SM00038; COLFI; 1.

DR SMART: SM00214; VWC; 1.

DR PROSITE: PS01208; VWFC_1; 1.

DR PROSITE: PS0184; VWFC_2; 1.

KW Collagen.

QY QY 1 GPPGPPGPTGLPPPG-----RGPGSGRFGPCADGVAGPKGAGERGSPGA 48

Db 180 GPPCAFPQGFQPPGPPGPPGSGGPMGRKPPGPPGKNGDDGAKYGRPPGPPGPPQ 239

QY 49 GPKGSPCEAGRPP-----GEAGLPCAAGLTGSPGSGPDGKTGPPGPAQD 93

Query Match 62.4%; Score 2176.5; DB 2; Length 1225;

Best Local Similarity 56.7%; Pred. No. 8.2e-79;

Matches 431; Conservative 28; Mismatches 157; Indels 144; Gaps 13;

Db 240 GARGLPAGLPGMKHGRFSGLDGAKGADGAPGKGPSPGNGAPQMGPRGLPGER 299
 Qy 94 GRPGPPGPPGARGQAGVGMFPKPKGAAGE-----PKKAGERGVPPGPGVAGPKDGEA 147
 Db 300 GRPGPPGTAGARGNDGAVGAGPPPTGPTGPPGPPGAVGAKGAGPQGARGSEGPGQVVR 359
 Qy 148 QAQPPGPPGAGPAGRGQGP-----AGSPGFQGLPGPAGPPGAGKPFGEQ 192
 Db 360 GEPGPPGPPGAGAGPAGNPGADQPGKAGKANGAPGAGPAGPAGPAGPAGPAGPAGP 419
 Qy 193 GVPDGLGAP-----GPSGA-----GEPGPTGLPGLPGRGPP 225
 Db 420 GNSGEPGAPKNGDGTAKGPFATGVQVPPGAGEEGKRGARGEPGSLPGLPGRGPP 479
 Qy 226 GSRGFPAGDVGAGPKGPPAGRGSPGAPKPSGCEAGLPGAKGLTGSPPSGPD 285
 Db 480 GSRGFPAGDVGAGPKGPPAGRGSPGAPKPSGCEAGLPGAKGLTGSPPSGPD 539
 Qy 286 GKTGPPGPPGAGDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 345
 Db 540 GKTGPPGPPGAGDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 599
 Qy 346 GKDEAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 405
 Db 600 GKDEAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 659
 Qy 406 GPSGPPAGE-----PGP-----TGLPGLPGRGPPGPPGPP 435
 Db 660 GPSGARGERFPGERGVQGPFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 719
 Qy 436 GAGVAGPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 486
 Db 720 GERGAAGLPKPKDGRDAGPKGADGSPGQGLPGPAGPAGPAGPAGPAGPAGPAGP 779
 Qy 487 GPDGKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 546
 Db 780 GPTGARGAFDREAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 839
 Qy 547 GPAGKQGEAGAQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 588
 Db 840 GPAGARGPAGPQGPGRDQKGTGQDGRGKGRHGRFSGLOGPPGPPGPPGPPGPPGPP 899
 Qy 589 ---GPGEAGKPCPE---QGVGPDGLGAGPP-----SGPAG 616
 Db 900 GPRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 939

RESULT 10

CALL CHICK
 ID CALL CHICK STANDARD; PRT; 1453 AA.
 AC P02457;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN Names=COL1A1;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE OF 1-153 FROM N.A.
 RX MEDLINE=88056316; PubMed=3678834; DOI=10.1016/0378-1119(87)90159-4;
 RA Finer M.H., Boedtker H., Doty P.;
 RT "Construction and characterization of cDNA clones encoding the 5' end
 of the chicken pro alpha 1(I) collagen mRNA.";
 RL Gene 56:71-78(1987).
 RN [2]
 RP SEQUENCE OF 1-144 FROM N.A.
 RX MEDLINE=88007542; PubMed=2820966;
 RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;

"Unusual DNA sequences located within the promoter region and the first intron of the chicken pro-alpha 1(I) collagen gene.";
 J. Biol. Chem. 262:13323-13332(1987).
 [3]

SEQUENCE OF 152-1187.
 RX MEDLINE=82231995; PubMed=7093229;
 RA Highbarger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H.,
 Gross J.;
 RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the complete primary structure of the helical portion of the chick skin collagen alpha 1(I) chain.";
 RL Biochemistry 21:2048-2055(1982).
 [4]

SEQUENCE OF 1200-1205.
 RX MEDLINE=72243016; PubMed=5047697;
 RA Eyre D.R., Glimcher M.J.;
 RT "Evidence for a previously undetected sequence at the carboxyterminus of the alpha 1 chain of chicken bone collagen.";
 RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
 [5]

SEQUENCE OF 981-1453 FROM N.A.
 RX MEDLINE=81160715; PubMed=6927845;
 RA Fuller F., Boedtker H.;
 RT "Sequence determination and analysis of the 3' region of chicken pro-alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids including the carboxy-terminal propeptide sequences.";
 RL Biochemistry 20:996-1006(1981).
 [6]

SEQUENCE OF 1311-1453 FROM N.A.
 RX MEDLINE=80134546; PubMed=6987088; DOI=10.1016/0014-5793(80)80761-7;
 RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
 Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
 RT "Nucleotide sequence of a collagen cDNA-fragment coding for the carboxyl end of pro alpha 1(I)-chains.";
 RL FEBS Lett. 111:61-65(1980).

-!- FUNCTION: Type I collagen is a member of group I collagen (fibrillar forming collagen).

-!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.

-!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and bones. In bones the fibrils are mineralized with calcium hydroxyapatite.

-!- PTM: Proline residues at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains. Pro-1153 is the only 3-hydroxypro and the only hydroxylated proline in position X.

-!- SIMILARITY: Belongs to the fibrillar collagen family.

-!- SIMILARITY: Contains 1 VWF domain.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M17839; AAA48704.1; -
 EMBL; M17838; AAA48704.1; JOINED.
 EMBL; V00401; CAA23695.1; -
 EMBL; M10571; AAA48671.1; ALT_SEQ.
 EMBL; M17607; AAA48672.1; -
 PIR; A27179; A27179.
 PIR; I50629; I50629.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR010007; VWF_C.
 Pfam; PF01410; COLFI; 1.
 Pfam; PF01391; Collagen; 18.
 Pfam; PF00093; VWC; 1.
 ProDom; PD000007; Clg_helix; 2.
 ProDom; PD002078; Fib_collagen_C; 1.
 SMART; SM00038; COLFI; 1.


```
DR SMART: SM00214; VWC: 1.
DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS50184; VWC_2; 1.
KW collagen; Direct protein sequencing; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;
KW Signal; Structural protein.
FT SIGNAL 1 22
FT PROPEP 23 151 N-terminal propeptide.
FT CHAIN 152 1205 Collagen alpha 1(I) chain.
FT PROPEP 1206 1453 C-terminal propeptide.
FT DOMAIN 31 89 VWC.
FT MOD_RES 152 152 Pyrrolidone carboxylic acid.
FT MOD_RES 160 160 Allene (By similarity).
FT MOD_RES 254 254 5-hydroxylysine (By similarity).
FT MOD_RES 851 851 5-hydroxylysine (Potential).
FT MOD_RES 1081 1081 5-hydroxyproline (Potential).
FT MOD_RES 1097 1097 5-hydroxylysine (Potential).
FT MOD_RES 1153 1153 3-hydroxyproline.
FT CARBOHYD 254 254 O-linked (Gal... ) (By similarity).
FT CARBOHYD 1354 1354 N-linked (GlcNAc... ) (By similarity).
FT CONFLICT 1187 1187 F -> L (in Ref. 5).
FT CONFLICT 1441 1441 Q -> H (in Ref. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Query Match 61.7%; Score 2151.5; DB 1; Length 1453;
Best Local Similarity 57.8%; Pred. No. 8.8e-78;
Matches 424; Conservative 17; Mismatches 175; Indels 117; Gaps 10;

QY 1 GPP-----GEPTGLPGPPGGRGSGRPPGADGVAGPKGPPAGERGSGPPA 48
DB 447 GPPGAGEGKRGARGEPGAGLPGPAGERGAPSRGPPGADGAGLPGKPPGERGSGAV 506
QY 49 GPKGSPGAGRGEAGLPGAKGLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQA 108
DB 507 GPKGSPGAGRGEAGLPGAKGLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQA 566
QY 109 GVMGPPGKGAAGEPKAGRGVCPGAVGPKGDKGAGAGQPPGAGPAGERGEGQA 168
DB 567 GVMGPPGKGAAGEPKAGRGVCPGAVGPKGDKGAGAGQPPGAGPAGERGEGQA 626
QY 169 GSPGFGGLPGPAGPGEAGKPGEQVCPDLGAPGSGPAGE-----PQP--- 212
DB 627 GAGFGGLPGPAGPGEAGKPGEQVCPDLGAPGSGPAGE-----PQP--- 686
QY 213 -----TGLPGPPGGRGSGRPPGADGVAGPKGPPAGERGSGPP----- 251
DB 687 GAGAPGNDGAKGDAGAPGAPGNEGPPGLEMGPPGGAAGLPGAKGDRGDPGPKGADGAP 746
QY 252 -----AGKPSGEGAGRP---GEAGLPGAKGLTSGSPGPDGKTGTPGGA----- 294
DB 747 GKDLGLRLGTIPGPPGAPAGCDKGEAGPPGPPGAGTARGAPGDRGEPGPPGAGFAGPP 806
QY 295 -----GQDRGPPPPGPPGARGOAGVWGPPGPKGAGEPKAGRGVVP 336
DB 807 GADQPGAKGTGDAKAGDAGPPGAGTGAAGPAGZVGPAGPPKAGSGAPGATGFP 866
QY 337 GPPGAVGPDGKDGAAGQPPGAPGAGE-----RGEQGPAGSPGFGGLPGAPGAGEAGKP 393
DB 867 GAAGRVGPPGSGNIGLPGPPGAPGKZSGKPRGETGPPAGRGPPGPPGPPGPKGSP 926
QY 394 GEQGVPGDLGAPGSPGAGEPGTGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 453
DB 927 GADGPIGAPGTPGPGIAGQVWGLPGQORGERGPPGLPGSPGEPKQSPGASGERGPP 986
QY 454 GPAGP---KSGPGEAGRGEAGLPGAKGLTSGSPGPDGKTGTPGAGQDGRPPGPP 510
DB 987 GPMGPPGLAGPGEAGREGAPGAEAPGRDGAAGPKGDRGETGPPGPPGAPGAPGVP 1046
QY 511 GARGOAGVWGPPGPKGAAGEPKKAGRGVPPGPPGAVGPPAGKDGE-----AGAQ 558
DB 1047 GPAGKNGRGETGPPGAPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1106
QY 559 GPP-----GPAGPAGERGEGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 603
```

```
DB 1107 GPPGPPGAPGEGQPSGASGAPGPPGSPGAGAGKGLGSLGPPGPPGRTGEGVP 1166
QY 604 GDLGAPGSPGAPG 616
DB 1167 GPPGPPGPPGPPG 1179

RESULT 11
QYIYB4
ID QYIYB4 PRELIMINARY; PRT; 1450 AA.
AC QYIYB4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
EX DOI=10.1002/(SICI)1097-0177(199909)216:1<59::AID-DVDY8>3.3.CO;2-2;
RA Ashina K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Clg helix.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clg helix; 4.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWC_1; UNKNOWN_1.
DR PROSITE; PS50184; VWC_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;

Query Match 58.6%; Score 2043.5; DB 2; Length 1450;
Best Local Similarity 54.7%; Pred. No. 1.6e-73;
Matches 401; Conservative 31; Mismatches 184; Indels 117; Gaps 11;

QY 1 GPP-----GEPGTLPGPPGGRGSGRPPGADGVAGPKGPPAGERGSGPPA 48
DB 444 GPPGSGEGRGSRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 503
QY 49 GPKGSPGAGRPPGAGLPGAKGLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQA 108
DB 504 GPKGSGESRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 563
QY 109 GVMGPPGKGAAGEPKAGRGVCPGAVGPKGDKGAGAGQPPGPPGPPGPPGPPGPP 168
DB 564 GVMGPPGKGAAGEPKAGRGVCPGAVGPKGDKGAGAGQPPGPPGPPGPPGPPGPP 623
QY 169 GSPGFGGLPGPAGPGEAGKPGEQVCPDLGAPGSPPAGE-----PGPTGLPGP- 218
DB 624 GSPGFGGLPGPAGPGEAGKPGEQVCPDLGAPGSPPAGE-----PGPTGLPGP- 683
QY 219 -----PGERGSGRGP-----PGERGSGRGP----- 231
DB 684 GSPGSPNDGAKGAGAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 743
```


QY	232	---	GADGVAGPKG	PAGERGSP-----GPAGPKSGPEAGRPGPAGLPGAKGLTGPS	279	
Db	744	KDKCARGLTGPT	GGPPGSGACGDKKEGGPSGPAFTGARSGPERGEFPAPGPAICGPP	803		
QY	280	GS----	RQPCKTGPPG	PACDQRPPPPPARGQAQVMGFPGFGAAAGEPKKAGERGVP	336	
Db	804	GADQGPGA	GESDGAFKGDAGAPGAPTGA	PGAVGNVAGPKGTGRGAAGPPCATGPF	863	
QY	337	GPPGAVGP	AGKDGEAGAQQPPGAPGACGERGECQGPAGSPGFCGLLP	PAGPP--GEAKP	393	
Db	864	GAAGRLGPP	PSGNAGPPGPPCGCKEKA	KSRETCTPARSRSEPPGAPP	PPSGEKKGSP	923
QY	394	GEOGVPRDL	GAPGSGPAGEPPTGLPDP	PEERGGPSRFFPGADVAGPKG	PAGERGSP	453
Db	924	GSDGPAGAPI	FPGQIGAQGVVLFPQ	RGERSGLPGPAGEDPKQPSFNGERGPP	983	
QY	454	GPAGP-----	KGSPGEAGRPCEAGLPGAK-----GLTSGPSGPDGKTGPP	495		
Db	984	GPSGPPCLGPP	GPGEPRGSGPSGEGAFRDGSGFKGDRGENGFSPPGAPGAPGAPV	1043		
QY	496	GPAGQD---	GRPCCPGPCARGQAQVMGFPGFGAAAGEPKAGE-----RGVPGPP	543		
Db	1044	GPAGKNXDRGET	GPAGPAGPAGPSGVRGAPGPACAGDKGEAGEQGERGMKHGRFNMQ	1103		
QY	544	GAUGPAGKDX	GEAGAQQPPGAPGAPAGERGEOGPAGSPGFCGLP	CPAGPPGEEAGKPCEEQVPP	603	
Db	1104	GPFGPPSSG	ECAFGPSGAPGPPPGSSGSTGDKGVNGLPPTGPPGPRGRNDVGPA	1163		
QY	604	GDLGAPGSP	GAPG	616		
Db	1164	GPFGPPGPPGPPG	1176			

RESULT 12
093251

ID	Q93251	PRELIMINARY;	PRT; 1445 AA.
AC	O93251;		
AD	01-NOV-1998	(TrEMBLrel. 08, Created)	
DT	01-NOV-1998	(TrEMBLrel. 08, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DT	Alpha 1 type I collagen.		
GN	Name=alpha 1 type I collagen;		
OS	Rana catesbeiana (Bull frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.		
OX	NCBI_TaxID=8400;		
RN	[1]		
SEQUENCE FROM N.A.			
RP	MEDLINE=99294154; PubMed=10367734; DOI=10.1016/S0945-053X(99)00005-0;		
RP	Asahina K., Utoh R., Obara M., Yoshizato K.;		
RT	"Cell-type specific and thyroid hormone-dependent expression of genes		
RT	of $\alpha 1(I)$ and $\alpha 2(I)$ collagen in intestine during		
RT	amphibian metamorphosis."		
RL	Matrix Biol. 18:89-103(1999).		
DR	EMBL; AB015440; BRA29028.1; -.		
DR	GO; GO:0005581; C:collagen; IEA.		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.		
DR	GO; GO:0006817; P:phosphate transport; IEA.		
DR	InterPro; IPRO08161; Clg.helix.		
DR	InterPro; IPRO08160; Collagen.		
DR	InterPro; IPRO0885; Fib_collagen_C.		
DR	InterPro; IPRO09041; PMP_SGCI.		
DR	InterPro; IPRO01007; VWF_C.		
DR	Pfam; PF01410; COLFI; 1.		
DR	Pfam; PF01391; Collagen; 18.		
DR	ProDom; PD000007; Clg.helix; 4.		
DR	ProDom; PD002078; Fib_collagen_C; 1.		
DR	SMART; SM00038; COLFI; 1.		
DR	SMART; SM00214; VWC; 1.		
DR	PROSITE; PS01208; VWF_C; 1.		
DR	PROSITE; PS0184; VWF_C; 1.		

[illegible]

RESULT 13	
Q640B2	
ID	Q640H
AC	Q640H
DT	25-00

Qy	340	GA	VGPAGKAGDGEAGACQPPGAPGAGE	-----RGEQGPAGSPGQFGLPGPAGPPGAGKGP	393
Db	864	GA	TRGVYGPFGPAGNSGPPGSPGAGKAGPRGETGPAGRPGEPGAAGPAGPPGKGGSP	923	
Qy	394	GE	QVPGDGLGACPGSPGAPGEPGGTGLPGPPGRRGGPGSGFGPCADGVAGPGKGPAGRGSP	453	
Db	924	GS	DGPAGAGFPIFGPGVAGSRGTIVGLPGMRGGRGFSGLPGPVGEPKQGPSPGSGRGPP	983	
Qy	454	GP	AGPAGP-----KSGPGEAGR-----PGBAGLPGAKGLT-----	492	
Db	984	GP	SGPPGLAGPPGCEAGREGAPGSEGAPGRDGAAGPKGRDGGGPGAGPPGAPGAPGAPGPV	1043	
Qy	493	GP	PCPAGQDGRGPPGPPGARGOAGVWGPGPKGAAGEPCKAGE-----	543	
Db	1044	GP	AGKSGDGRGTGPAGPAGPAGVAGARGPAGPQGPGRDKGAGEQGERGMKGRHGFNGPA	1103	
Qy	544	GA	VGPAGKAGDGEAGACQPPGAPGAPGAPGAGRGEGPAGSPGQFGLPGPAGPPGAGKGPGEQGV	603	
Db	1104	GP	PCPPGSGEGQPGSCSAGPAGPRGPPGSGNPKDGANGLPGPIGPPGPRGRTGDVGA	1163	
Qy	604	GD	LGAFGPSPGAPG	616	
Db	1164	GP	PPGPPGPPGPPG	1176	

RESULT 14

Q802B5	PRELIMINARY;	PRT;	1449	AA.	
ID	Q802B5	PRELIMINARY;	PRT;	1449	AA.
AC	Q802B5;				
DT	01-JUN-2003	(TEMBLrel. 24, Created)			
DT	01-JUN-2003	(TEMBLrel. 24, Last sequence update)			
DT	01-MAR-2004	(TEMBLrel. 26, Last annotation update)			
DE	Collal-prov protein.				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	SEQUENCE FROM N.A.				
RP	TISSUE=Embryo;				
RC	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Diatchenko L., Marsden K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McSwann P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,				
RA	Krzywinski M.I., Skalka U., Smaluk D.E., Schnerch A., Schein J.B.,				
RA	Jones S.J., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	TISSUE=Embryo;				
RC	MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;				
RA	Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,				
RA	Richardson P.;				
RT	"Genetic and genomic tools for Xenopus research: The NIH Xenopus				
RT	Dev. Dyn. 225:384-391(2002).				
RL	[3]				
RN	SEQUENCE FROM N.A.				


```

QY 583 GLPDPAGPPBAGKPGEGVPGDLGAPGSPGAG 616
DB 1098 GFNGPSPGPPGPGSGEGQSPGASGAPGRGPPG 1131

```

RESULT 15	CALL_RAT	STANDARD;	PRT;	671 AA.
AC	P02454; P02455;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	Collagen alpha 1(I) chain (Fragments).			
GN	Name=Collal;			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE OF 1-19.			
RX	MEDLINE=69155173; PubMed=5777344;			
RA	Bornstein P.;			
RT	"Comparative sequence studies of rat skin and tendon collagen. II. The			
RT	absence of a short sequence at the amino terminus of the skin alpha-1			
RT	chain.";			
RL	Biochemistry 8:63-71(1969).			
RN	[2]			
RP	SEQUENCE OF 5-19.			
RX	MEDLINE=67162268; PubMed=5337886;			
RA	Kang A.H., Bornstein P., Piez K.A.;			
RT	"The amino acid sequence of peptides from the cross-linking region of			
RT	rat skin collagen.";			
RL	Biochemistry 6:788-795(1967).			
RN	[3]			
RP	SEQUENCE OF 20-55.			
RX	MEDLINE=67165368; PubMed=4290711;			
RA	Bornstein P.;			
RT	"The incomplete hydroxylation of individual prolyl residues in			
RT	collagen.";			
RL	J. Biol. Chem. 242:2572-2574(1967).			
RN	[4]			
RP	SEQUENCE OF 56-102.			
RX	MEDLINE=71263178; PubMed=4327399;			
RA	Butler W.T., Ponds S.L.;			
RT	"Chemical studies on the cyanogen bromide peptides of rat skin			
RT	collagen. Amino acid sequence of alpha 1-CB4.";			
RL	Biochemistry 10:2076-2081(1971).			
RN	[5]			
RP	SEQUENCE OF 103-139.			
RX	MEDLINE=70085124; PubMed=5411206;			
RA	Butler W.T.;			
RT	"Chemical studies on the cyanogen bromide peptides of rat skin			
RT	collagen. The covalent structure of alpha 1-CB5, the major hexose-			
RL	containing cyanogen bromide peptide of alpha 1.";			
RL	Biochemistry 9:44-50(1970).			
RN	[6]			
RP	SEQUENCE OF 140-238.			
RX	MEDLINE=72136131; PubMed=4335087;			
RA	Ballan G., Click E.M., Bornstein P.;			
RT	"Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of			
RT	the hydroxylamine-produced fragment HA1.";			
RL	Biochemistry 10:4470-4478(1971).			
RN	[7]			
RP	SEQUENCE OF 239-418.			
RX	MEDLINE=73006942; PubMed=4342027;			
RA	Ballan G., Click E.M., Hermodson M.A., Bornstein P.;			
RT	"Structure of rat skin collagen alpha 1-CBB. Amino acid sequence of			
RT	the hydroxyl amine-produced fragment HA2.";			
RL	Biochemistry 11:3798-3806(1972).			
RN	[8]			
RP	SEQUENCE OF 419-567.			
RX	MEDLINE=74271984; PubMed=4366532;			
RA	Butler W.T., Underwood S.P., Finch J.E. Jr.;			

Search completed: June 17, 2005, 15:13:51
Job time : 96.5221 secs

F I	SALT-EXTRACTED RAT SKIN C
SQ	SEQUENCE
	671 AA:
	60615 MW:
	9DC3114204AC4918 CRC64:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 134.148 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGEPPTGLPGRGGRGG.....GEQVPGDLGAPGSPGAG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_l6Dec04.*

- 1: geneseq1980s.*
- 2: geneseq1990s.*
- 3: geneseq2000s.*
- 4: geneseq2001s.*
- 5: geneseq2002s.*
- 6: geneseq2003as.*
- 7: geneseq2003bs.*
- 8: geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4640	100.0	821	ADM48392	Adm48392 Human rec
2	3482	75.0	617	ADM48391	Adm48391 Human rec
3	3482	75.0	617	ADQ26217	Adq26217 Human gel
4	2765	59.6	1449	AAE02535	Aae02535 Porcine a
5	2765	59.6	1057	AAy84544	Aay84544 A human c
6	2765	59.6	1058	AAy84403	Aay84403 Amino aci
7	2765	59.6	1107	AAy84540	Aay84540 Amino aci
8	2765	59.6	1171	AAy84538	Aay84538 A chimeri
9	2765	59.6	1464	AAW68485	Aaw68485 Human rec
10	2765	59.6	1464	ADD45059	Add45059 Human pro
11	2765	59.6	1464	ADD45055	Add45055 Human pro
12	2765	59.6	1464	ADD45051	Add45051 Human pro
13	2765	59.6	1464	ADR16800	Adri6800 Human col
14	2765	59.6	1464	ADR16425	Adri6425 Human col
15	2765	59.6	1464	ADR99144	Adr99144 Collagen,
16	2761	59.5	1461	ABg93947	Abg93947 Human pol
17	2761	59.5	1464	AAU14136	Aau14136 Human nov
18	2759	59.5	1057	AAy84541	Aay84541 Amino aci
19	2757	59.4	1388	AAy84539	Aay84539 Amino aci
20	2755	59.4	1464	AAAB2454	Aab2454 Human pro
21	2755	59.4	1464	ABB90764	Abb90764 Human tum
22	2755	59.4	1464	ABP68610	Abp68610 Human pan
23	2755	59.4	1464	ABU54471	Abu54471 Human tum
24	2755	59.4	1464	ABr47417	AbR47417 Breast ca
25	2755	59.4	1464	ABR92064	ABr92064 Human cer

26	2755	59.4	1464	7	ADD14142	Add14142 Human src
27	2755	59.4	1464	7	ADP65246	Adp65246 Human alp
28	2755	59.4	1464	8	ADQ19470	Adq19470 Human sof
29	2755	59.4	1464	8	ADQ29653	Adq29653 Human col
30	2752.5	59.3	1463	4	AAE02532	Aae02532 Bovine al
31	2752	59.3	1464	7	ADe87048	Ade87048 Human pan
32	2750	59.3	1107	2	AAR89472	Aar89472 Collagen/
33	2750	59.3	1169	2	AAR89469	Aar89469 Collagen/
34	2750	59.3	1169	2	AAy84537	Aay84537 Amino aci
35	2750	59.3	1171	2	AAR89470	Aar89470 Collagen/
36	2750	59.3	1388	2	AAy89471	Aay89471 Collagen/
37	2748.5	59.2	1453	7	ADD45053	Add45053 Rat Prote
38	2748.5	59.2	1453	7	ADD45057	Add45057 Rat Prote
39	2748.5	59.2	1453	7	ADD48341	Add48341 Rat Prote
40	2748.5	59.2	1453	7	ADD45049	Add45049 Rat Prote
41	2748.5	59.2	1453	7	ADD48337	Add48337 Rat Prote
42	2748.5	59.2	1453	7	ADD48345	Add48345 Rat Prote
43	2732.5	58.9	1453	5	ABG93948	Abg93948 Mouse pol
44	2725	58.7	1536	7	ADe87051	Ade87051 Human pan
45	2723.5	58.7	1411	3	AAy56800	Aay56800 Human pre

ALIGNMENTS

RESULT 1

ADM48392
ID ADM48392 standard; protein; 821 AA.

XX AC ADM48392;

XX 03-JUN-2004 (first entry)

XX Human recombinant gelatin-like polypeptide Hu-4.

XX Plasma substitute; Gelatin-like protein; plasma expander; human.

XX Homo sapiens.

XX EPI398324-A1.

XX 17-MAR-2004.

XX 11-SEP-2002; 2002EP-00078745.

XX 11-SEP-2002; 2002EP-00078745.

XX (FUJF) FUJI PHOTO FILM BV.

XX Bouwstra JB, Toda Y;

XX WPI; 2004-229415/22.

Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.

Example 1; SEQ ID NO 3; 31pp; English.

The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a tetramer of human recombinant gelatin-like plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgE antibodies. The gelatin-like proteins can be covalently attached to

CC pharmaceutically active compounds. After administration, the coupled
 CC medicament will not diffuse from the circulating blood into the
 CC interstitium. Clearance by liver and kidney will be kept to a minimum,
 CC ensuring a more constant plasma level of the medicament. Suitable
 CC medicaments include those involved in intervening blood clotting,
 CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,
 CC immune responses, and blood levels of messenger molecules such as
 CC hormones.
 XX
 SQ Sequence 821 AA;
 Query Match 100.0%; Score 4640; DB 8; Length 821;
 Best Local Similarity 100.0%; Pred. No. 3.2e-278;
 Matches 821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 60
 DB 1 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 60
 QY 61 GEAGLPGAKGLTSGPFGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
 DB 61 GEAGLPGAKGLTSGPFGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
 QY 121 GEPKAGGERGVPPGAVGAGKDEAGAGPAGPAGPAGERGSGPAGKSGPGEAGRP 180
 DB 121 GEPKAGGERGVPPGAVGAGKDEAGAGPAGPAGPAGERGSGPAGKSGPGEAGRP 180
 QY 181 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 240
 DB 181 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 240
 QY 241 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 300
 DB 241 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 300
 QY 301 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 360
 DB 301 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 360
 QY 361 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 420
 DB 361 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 420
 QY 421 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 480
 DB 421 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 480
 QY 481 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 540
 DB 481 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 540
 QY 541 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 600
 DB 541 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 600
 QY 601 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 660
 DB 601 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 660
 QY 661 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 720
 DB 661 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 720
 QY 721 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 780
 DB 721 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 780
 QY 781 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 821
 DB 781 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 821

RESULT 2

ADM48391
 ID ADM48391 standard; protein; 617 AA.
 XX AC ADM48391;
 XX DT 03-JUN-2004 (first entry)
 XX DE Human recombinant gelatin-like polypeptide Hu-3.
 XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.
 XX OS Homo sapiens.
 XX PN EP1398324-A1.
 XX PD 17-MAR-2004.
 XX PF 11-SEP-2002; 2002EP-00078745.
 XX PR 11-SEP-2002; 2002EP-00078745.
 XX PA (FUJF) FUJI PHOTO FILM BV.
 XX PI Bouwstra JB, Toda Y;
 XX DR WPI; 2004-229415/22.
 XX PS Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.
 XX PT Example 1; SEQ ID NO 2; 31pp; English.
 CC The present sequence is the protein sequence of recombinant gelatin-like
 CC protein Hu-3. This is a trimer of human recombinant gelatin-like protein
 CC Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma
 CC comprises a solution of saline and a protein having a colloid osmotic
 CC function. The protein is a recombinant gelatin-like protein with a
 CC molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric
 CC point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The
 CC recombinant gelatin-like protein, or a dimer, trimer or tetramer of the
 CC protein, is useful as a plasma expander that has a lower clearance rate
 CC from blood circulation, provides better and predictable regulation of
 CC clearance rate and which is less susceptible to proteolytic degradation
 CC than presently used gelatin derivatives. Recombinant gelatin-like
 CC proteins that are in essence free of hydroxyproline do not give rise to
 CC an immunological reaction with blood samples containing IGE antibodies.
 CC The gelatin-like proteins can be covalently attached to pharmaceutically
 CC active compounds. After administration, the coupled medicament will not
 CC diffuse from the circulating blood into the interstitium. Clearance by
 CC liver and kidney will be kept to a minimum, ensuring a more constant
 CC plasma level of the medicament. Suitable medicaments include those
 CC involved in intervening blood clotting, vasodilation, function of
 CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood
 CC levels of messenger molecules such as hormones.
 XX SQ Sequence 617 AA;
 Query Match 75.0%; Score 3482; DB 8; Length 617;
 Best Local Similarity 100.0%; Pred. No. 5.4e-207;
 Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 60
 DB 1 GPPGEPGTLGPPGGERGSGRFGADGAGVAGPKPAGERGSGPAGKSGPGEAGRP 60
 QY 61 GEAGLPGAKGLTSGPFGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
 DB 61 GEAGLPGAKGLTSGPFGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
 QY 121 GEPKAGGERGVPPGAVGAGKDEAGAGPAGPAGPAGERGSGPAGKSGPGEAGRP 180
 DB 121 GEPKAGGERGVPPGAVGAGKDEAGAGPAGPAGPAGERGSGPAGKSGPGEAGRP 180

QY 181 GPPGAGKPGGQVGDGLGAPGSPGAGEPCTGLPGPPGRRGGSPGRRPGADGVAGPK 240
 DB 181 GPPGAGKPGGQVGDGLGAPGSPGAGEPCTGLPGPPGRRGGSPGRRPGADGVAGPK 240
 QY 241 GPAGERSGPGAGKSGSGEAGRFCEAGLPKAGLGTSGSPGPDGKTGTPPGAGQDGRP 300
 DB 241 GPAGERSGPGAGKSGSGEAGRFCEAGLPKAGLGTSGSPGPDGKTGTPPGAGQDGRP 300
 QY 301 GPPGPPGARGQAGVMGPPGPKGAGEPKKAGERGVPVPPGAVGPKDGEAGAGQPPGPA 360
 DB 301 GPPGPPGARGQAGVMGPPGPKGAGEPKKAGERGVPVPPGAVGPKDGEAGAGQPPGPA 360
 QY 361 GPAGERGQSPAGSPGFGGLPGPAGPCEAGKPGEGVPPGDLGAPGSGPAGEPGPTGLP 420
 DB 361 GPAGERGQSPAGSPGFGGLPGPAGPCEAGKPGEGVPPGDLGAPGSGPAGEPGPTGLP 420
 QY 421 GPPGERGSGRGGFPDAGVAGPKGAGERSGPPAGPKGSPGAGRFGAGLPKAGLGT 480
 DB 421 GPPGERGSGRGGFPDAGVAGPKGAGERSGPPAGPKGSPGAGRFGAGLPKAGLGT 480
 QY 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVMGPPGKGAAGEPKKAGERGVP 540
 DB 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVMGPPGKGAAGEPKKAGERGVP 540
 QY 541 GPPGAVGPKDGEAGAGQPPGPPGAGPAGERGEGQSPGSPGFGGLPGPAGPPGEGAGKPGEQ 600
 DB 541 GPPGAVGPKDGEAGAGQPPGPPGAGPAGERGEGQSPGSPGFGGLPGPAGPPGEGAGKPGEQ 600
 QY 601 GVPDGLGAPGSPGAG 616
 DB 601 GVPDGLGAPGSPGAG 616

RESULT 3

ADQ26217
 ID ADQ26217 standard; protein; 617 AA.
 XX
 AC ADQ26217;
 DT 23-SEP-2004 (first entry)
 XX
 DE Human gelatine-like polypeptide Hu-3.
 XX
 KW Human; gelatine-like protein; Hu-3; microcarrier; cell culture.
 XX
 OS Homo sapiens.
 XX
 PN WO2004056976-A2.
 XX
 PD 08-JUL-2004.
 XX
 PF 23-DEC-2003; 2003WO-NL000922.
 XX
 PR 23-DEC-2002; 2002EP-00080539.
 XX
 PA (FUJF) FUJI PHOTO FILM BV.
 XX
 PI Bouwstra JB, Van Es AJJ, Toda Y;
 XX
 DR WPI; 2004-507711/48.
 XX
 PT Preparing cell culture support useful for culturing anchorage dependent
 cells, involves coating microcarrier bead with gelatine or gelatine-like
 PT protein having specific molecular weight.
 XX
 PS Example 1; SEQ ID NO 1; 19pp; English.

CC The present sequence is the protein sequence of human recombinant gelatin
 CC -like polypeptide Hu-3, which has a molecular weight of approximately 54
 CC kDa and which contains multiple Gly-Xaa-Yaa triplets. In an example from
 CC the invention, recombinant Hu-3 was immobilised on polystyrene beads
 CC using a heterobifunctional crosslinking agent, and used in a spinner
 CC flask cell culture. A claimed process for the preparation of a cell

CC culture support comprises the step of coating a microcarrier bead with
 CC gelatine or gelatine-like protein having a molecular weight of about 40-
 CC 200 kDa, and optionally further comprising the step of immobilising the
 CC gelatine or gelatine-like protein on the microcarrier. In this process,
 CC more than 75%, preferably more than 85% and more preferably more than 95%
 CC of the gelatine or gelatine-like protein has the same molecular weight.
 CC The gelatine or gelatine-like protein is recombinantly produced to obtain
 CC a material of uniform molecular weight and to reduce the risk of
 CC contamination with prions. A claimed cell support consists of microbeads
 CC of 50-500 um size coated with a gelatine-like protein consisting of at
 CC least 95% Gly-Xaa-Yaa triplets and containing at least 15% proline
 CC residues and less than 5% of hydroxyproline residues, with a molecular
 CC weight distribution showing a maximum between 40 and 200 kDa, at least
 CC 75% of the protein molecules having a molecular weight within 2% of the
 CC maximum, large-scale production of expressed products can be accomplished
 CC with gelatine-coated microcarriers.

XX Sequence 617 AA;

Query Match 75.0%; Score 3482; DB 8; Length 617;
 Best Local Similarity 100.0%; Pred. No. 5.4e-207;
 Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPGPPGRRGGSPGRRPGADGVAGPKGAGERSGPPAGPKGSGEAGRP 60
 DB 1 GPPGEPGPTGLPGPPGRRGGSPGRRPGADGVAGPKGAGERSGPPAGPKGSGEAGRP 60
 QY 61 GEAGLPKAGLGTSGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVMGPPGPKGAA 120
 DB 61 GEAGLPKAGLGTSGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVMGPPGPKGAA 120
 QY 121 GEPKAGERGVPVPPGAVGPKDGEAGAGQPPGPPGAGPAGERGEGQSPGFGGLPGPA 180
 DB 121 GEPKAGERGVPVPPGAVGPKDGEAGAGQPPGPPGAGPAGERGEGQSPGFGGLPGPA 180
 QY 181 GPPGAGKPGEGQVPPGDLGAPGSPGAGEPCTGLPGPPGRRGGSPGRRPGADGVAGPK 240
 DB 181 GPPGAGKPGEGQVPPGDLGAPGSPGAGEPCTGLPGPPGRRGGSPGRRPGADGVAGPK 240
 QY 241 GPAGERSGPGAGPKSGSGEAGRFCEAGLPKAGLGTSGSPGPDGKTGTPPGAGQDGRP 300
 DB 241 GPAGERSGPGAGPKSGSGEAGRFCEAGLPKAGLGTSGSPGPDGKTGTPPGAGQDGRP 300
 QY 301 GPPGPPGARGQAGVMGPPGPKGAGEPKKAGERGVPVPPGAVGPKDGEAGAGQPPGPA 360
 DB 301 GPPGPPGARGQAGVMGPPGPKGAGEPKKAGERGVPVPPGAVGPKDGEAGAGQPPGPA 360
 QY 361 GPAGERGQSPAGSPGFGGLPGPAGPCEAGKPGEGVPPGDLGAPGSGPAGEPGPTGLP 420
 DB 361 GPAGERGQSPAGSPGFGGLPGPAGPCEAGKPGEGVPPGDLGAPGSGPAGEPGPTGLP 420
 QY 421 GPPGERGSGRGGFPDAGVAGPKGAGERSGPPAGPKGSPGAGRFGAGLPKAGLGT 480
 DB 421 GPPGERGSGRGGFPDAGVAGPKGAGERSGPPAGPKGSPGAGRFGAGLPKAGLGT 480
 QY 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVMGPPGKGAAGEPKKAGERGVP 540
 DB 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVMGPPGKGAAGEPKKAGERGVP 540
 QY 541 GPPGAVGPKDGEAGAGQPPGPPGAGPAGERGEGQSPGSPGFGGLPGPAGPPGEGAGKPGEQ 600
 DB 541 GPPGAVGPKDGEAGAGQPPGPPGAGPAGERGEGQSPGSPGFGGLPGPAGPPGEGAGKPGEQ 600
 QY 601 GVPDGLGAPGSPGAG 616
 DB 601 GVPDGLGAPGSPGAG 616

RESULT 4

AAE02535
 ID AAE02535 standard; protein; 1449 AA.
 XX
 AC AAE02535;

10-AUG-2001	(first entry)
Porcine alphas(I) collagen.	
Porcine alpha(I) collagen; gelatin; cytostatic; viral infection; pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine; medical; arterial sealant; bone graft; dermal implant; haemostat; cancer; rheumatoid arthritis; beverage; photographic application.	
Sus scrofa.	
Key	Location/Qualifiers
Misc-difference	829..830
/note=	"Encoded by ggcgaacctggtgatgctgtgctaaaggcgatctggtccccgcctctgga"
WO2001134647-A2.	
17-MAY-2001.	
10-NOV-2000; 2000WO-US030792.	
12-NOV-1999; 99US-00439058.	
10-NOV-2000; 2000US-00709700.	
(FIBR-) FIBROGEN INC.	
Bell MP, Neff TB, Polarek JW, Seeley TW; WPI; 2001-335911/35. N-PSDB; RAD06576.	
Novel isolated and purified bovine or porcine collagens and gelatins useful in medical, pharmaceutical, food and cosmetic industries, as vaccine, and for treating autoimmune disorders, infections and cancer.	
Example 3; Fig 8; 168pp; English.	
The present sequence is porcine alpha(I) collagen. The present invention relates to recombinant synthesis of collagens and gelatins derived from animals. Collagen is useful in medical, pharmaceutical, food and cosmetic industries. Collagen is an important component of arterial sealants, bone grafts, drug delivery system, dermal implants, haemostats, and incontinence implants, and for treating autoimmune disorders such as rheumatoid arthritis. Collagen is useful in food products such as sausage casings, and in cosmetics or facial and skin products such as sausage moisturisers. Recombinant gelatin is useful in vaccine formulations for treating viral infections, autoimmune diseases and cancer. Gelatin is useful in the manufacture or as a component of various pharmaceutical and medical devices and products, in food and beverage industries, in hair care and skin care products, as a glue or adhesive in various manufacturing processes, as a light-sensitive coating in various electronic devices, as photorealist base in photolithographic processes, in printing and photographic applications, in laboratory application, and as a component in various gels used for biochemical and electrophoretic analysis, including enzymographic gels	
Sequence 1449 AA;	

Db	313	GRPGPPGAGARGNDGATGAAGPGGPTGAGPPGFCAVGAKGEAGFCQAGSGEGPQGV	372
Qy	148	GAQGPFGPAGPACERGEQGF-----AGSPGQGLPGAPGPPGAGKPGQ	192
Db	373	GEFGPPGAGAGPAGNPGADQPGCKGANGAPGTAGAFGFCARGPSGPGSPGPPGK	432
Qy	193	GVPGDLDGAP-----GPSGA-----GEFGPTGLPGPPGERGGP	225
Db	433	GNSGEFGAGSGKGDGAKGEPGTVGQSPGAGBEGKRGARFPGPAGLPGPPGERGGP	492
Qy	226	GSRGFFGADGVAGPKPACERGSPPGAPKSGSPGAGRPGEAGLPCAKGLTQSPSGP	285
Db	493	GSRGFFGADGVAGPKPACERGSPPGAPKSGSPGAGRPGEAGLPCAKGLTQSPSGP	552
Qy	286	GKTGPPGAGQDGRPGPPGARGOAGVMGPPGPKGAAGFEGKAGERGVPPPGAVGA	345
Db	553	GKTGPPGAGQDGRPGPPGARGOAGVMGPPGPKGAAGFEGKAGERGVPPPGAVGA	612
Qy	346	GKDGAGAGQPPGAPGAGERGEQGPAGSPGQGLPGPAGRPGEAGKPGEOGVPGDLGAP	405
Db	613	GKDGAGAGQPPGAPGAGERGEQGPAGSPGQGLPGPAGRPGEAGKPGEOGVPGDLGAP	672
Qy	406	GPSGPGAGE-----PGP-----TGLPGPPGERGGPSGRGFP	435
Db	673	GPSGARGRGFPFGERGVQPPGPPGAPRGANGAPGNDGAKGADGAPGAGPSQAGPCLQMP	732
Qy	436	GADGVAGPKGAGERGSPPGAPKSGPGE-----AGRPGAGLPCAKGLTQSPSGP	486
Db	733	GERGAAGLPGPKDGRDAGPKGADGAPGQGVRLTGPDPGPPGAGPDEKGETGPGSPA	792
Qy	487	GDGKTGPPGAGQDGRPGP---PGPPGARGOAGVMGF---PGPKGAAGFEGKAGERGV	539
Db	793	GPTGARGAPGDRGFPFGPPGAGFAGPPGADQPGAKGPTGPPGPGISGVGAPGKARGS	852
Qy	540	PGPPCA-----VGPAGKDGAGAGQPPGAPGAGE---RGEQGPAGSPGQGLPGP	587
Db	853	AGPPGATGFPCAAGRVGFPGPSGNAGPPPPGPAKSGSKPRGETGAPRPGSAGPPG	912
Qy	588	AGPPGEAGKPGEOQGVPGDLGAPSGPAGBFGPTGLFPGPPGERGGPSGRFPAGDGVAGP	647
Db	913	PGPAGERGSPGADGAPAGTPTGPGQITAGQKVVGLPGQGERGFPCLPGPSGEPGQGP	972
Qy	648	KGPAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLTGSP---GSPGPD---GKTGPPGP	701
Db	973	SGPSGGERGPPGMPG---PGLAGPPGESGREGAPGABGSGPRGDGAPFKGDRGESGAPG	1029
Qy	702	AGQDGRPGPPGPGARQOAGVMGFPFGPKGAAGFEGKAGERGVPPGCAVGPAGKQGE---	758
Db	1030	PGAPGAPGAPGPPGAPGKSGDRGETGAPGAPGVPVVGARGPAGPPRGDKGETGEQGD	1089
Qy	759	-----AGAQQPPGP-----AGPAGERGEQGPAGSP---GFQGLPGPAGP	794
Db	1090	RGIKGHRGFSGLQGPFGPQSPGEQGPAGPRGPPGSGAGAPCKDGLNGLPGPIGP	1149
Qy	795	PGERAGKPGEOGVPGDLGAPGSPGAP	820
Db	1150	PGPRGRTGDAGPVGPPGPPGPPGPPG	1175

RESULT 5
 AAY84544
 ID AAY84544 standard; protein; 1057 AA.
 XX
 AC AAY84544;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE A human collagen 1 (alpha) protein helical region.
 XX
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation.

XX Homo sapiens.
 XX EP992586-A2.
 FN 12-APR-2000.
 PD 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 PR (USSU) US SURGICAL CORP.
 PA Gruskin EA, Buechter DD, Zhang G, Connolly K;
 PI WPI; 2000-259138/23.
 XX N-PSDB; AAA12503.
 DR Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX Example 10; Fig 39A-E; 260pp; English.
 PS The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents human collagen 1 (alpha1) helical region,
 CC which may be produced using the method of the invention
 SQ Sequence 1057 AA;

Query Match 59.6%; Score 2765; DB 3; Length 1057;
 Best Local Similarity 55.4%; Pred. No. 1.2e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGPGSGRFGPGADGVAGPKPAGRGSGPGAGPKGSGPGEA--- 57
 DB 18 GPMGSGRGLFPGPGAGPGQFQPGPEPGAGSGPMGRPGPGPKNGDDGEAGKP 77
 QY 58 GRPGAGLP---GAKGLTSGPSGP------DGKTGPPGAGQDGRPGPPGPGAR 105
 DB 78 GRPGRGPPGQAGRLPGLTGLPKMGRGFGSLDGAKDAGPAGPKGPGSGPENGAP 137
 QY 106 GQAGVMGFRGPKGAGEPKAGER-----GVPPGPGAVGA------GKDEAGAO 150
 DB 138 GQMGPRGLPGRGRPGPGAGARGNDGATGAAGPPGPTGPGPPGPGAVGAKGAGPQ 197
 QY 151 GPPGPGAGRGEGGPG-----AGSPGFQGLPGPA 180
 DB 198 GPRGSEGVGVRGEPGPGPAGAGPAGNPGADGQPGAKGANGAPGAGPFGPGARGPS 257
 QY 181 GPPGAGKPGGVGPDILGAP-----GPSGA-----GEPGPT 213
 DB 258 GPGGPGPPGKNGSGEPGAPGSGKDTGAKGPGPGVGVQVPGPGAGEGKRGARGEPGT 317
 QY 214 GLPGRGPGRGPGSGRFGADGVAGPKPAGRGSGPGAGPKGSGPGEAGRPGEAGLPKAK 273

DB 318 GLPGRGPGRGSGRFGADGVAGPKPAGRGSGPGAGPKGSGPGEAGRPGEAGLPKAK 377
 QY 274 GLTSGPSGPDGKTGTPPGAGQDGRPGPPGARGOAGVMGPPGPKGAAGEPGKAGER 333
 DB 378 GLTSGPSGPDGKTGTPPGAGQDGRPGPPGARGOAGVMGPPGPKGAAGEPGKAGER 437
 QY 334 GVPGRGAVGAGKDGAGAGAGGPGPAGPAGRGEGEQGPGAGSPGFGQLPGPAGPGEAGKP 393
 DB 438 GVPGRGAVGAGKDGAGAGAGGPGPAGPAGRGEGEQGPGAGSPGFGQLPGPAGPGEAGKP 497
 QY 394 GEOGVPGDLGAPGSPAGE-----PGP-----TGLPGPP 423
 DB 498 GEOGVPGDLGAPGSPGARGERFPGRGVQVQPPGAGPRGANGAPNDGAKGDAGAPGAP 557
 QY 424 GERGPGSRGPGADGVAGPKPAGRGSGPGAGPKGSGPGE-----AGRPGEAGLP 474
 DB 558 GSQAGPLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSGPKGDKVRLGTGTGTPGPGAPAG 617
 QY 475 GAK-----GLTSGPSGPDGKTGTPPGAGQDGRPGPPGPGARGO----- 515
 DB 618 GDKGESGPGAGPTGARGAPGDRGEPGPPGAGPAGPPGADGQPGAKGEPGDGAKGDA 677
 QY 516 -----AGVMGFRGPKGAAGEPKKAGERGVPPGPGA-----VGPAGKDGEGAGAOQPP 561
 DB 678 GPPGPGAGPDPGPIGNVGAAPGAKGARGSGAGPPGATGTPGAGRVGPPGPGSNAGPPGPP 737
 QY 562 GPAGP-----AGRGQAGSAGSPGFGQLPGPAGPAGRGAGKPGGQGVPGDLGAPGSPGAGEP 618
 DB 738 GPAGKEGKGRGTGTGTPAGRGVGGPPGPPGAGKEGSGADGAGPAGTGTGPGQIAGQR 797
 QY 619 GPTGLPGRGPGRGSGRFGADGVAGPKPAGRGSGPGAGP---KGSPPGAGRPGEA 675
 DB 798 GVVGLPQGRGPRGPPGLPGSGRPGKQGPSASGERGPPGPMGPPGLAGPPGEGSREGAP 857
 QY 676 GLPKAGLITSGPSGPDGKTGTPPGAGQDGRPGPPGARGOAGVMGFRGPKGAAGEP 735
 DB 858 GAESGPRGDSGAKGDRGTGTGTPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 917
 QY 736 GKAGERGVPPGPGAVGAPGAKDGE-----AGAQGPPG-----AQPA 771
 DB 918 GPAGARGPAGPQGRGDKGTGTEGQDRGKIGKHRGFGSLQGGPPGPGSGEQGPGSGASGA 977
 QY 772 GEREGQGAGSP---GFGQLPGPAGPGEAGKPGEQGVPGDLGAPGSGPAG 820
 DB 978 GPRGPGSAGAPGKDXGLNLGLPFGPPGPRGRTDAGVGPVPPGPPGPPG 1029

RESULT 6
 AAY84403
 ID AAY84403 standard; protein; 1058 AA.
 XX AAY84403;
 AC AAY84403;
 XX 12-JUL-2000 (first entry)
 DT Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
 DE Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
 XX collagen; mussel adhesive protein; bioadhesive.
 OS Homo sapiens.
 XX WO200014201-A1.
 FN 16-MAR-2000.
 PD 07-SEP-1999; 99WO-US020462.
 PF 09-SEP-1998; 98US-0099652P.
 PR (USSU) US SURGICAL CORP.
 PA (PAOL/) PROLELLA D N.
 PA (GRUS/) GRUSKIN E A.

The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily produce extracellular matrix proteins. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell, incorporating the nucleic acid sequence into the cell, and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibronectin and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a chimeric collagen 1 (alpha1)/decorin protein, which may be produced using the method of the invention

Query Match 59.6%; Score 2765; DB 3; Length 1107;
 Best Local Similarity 55.4%; Pred. No. 1.3e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

1 GPPGPGTGLPGRGPGSGRFGPGADGVAGPKGPGAGRGSPGPGPKGSPGCEA--- 57
 18 GPMGSPGRGLPGPGAPGPGQFGPGEPGPGASGPMGPRGPPGPKGNDGEGAKP 77
 58 GRPGAGLP---GAKGLTSGPSGPGP-----DGKTGPPGAGQGRGPPGPGAR 105
 78 GRPGGPGPGQAGRLPCTAGLPMKGRGSLDGAKGADGAGPAGPKGPGSPGNGAP 137
 106 GQAGVMGPPGPKGAAGEPKAGER-----GVPGPPGAVGPA-----GKDGAEAG 150
 138 GQMPRGLPGERGRGAGPAGARGNDGATGAGPPGTGAGPPGPGAVGAKGEGAPQ 197
 151 GPPGAGPAGERGEGP-----AGSPFGQLGPGA 180
 198 GPRGSEGGVGRGPGPGAGAGPAGNPGADGQPGAKGANGAPGAGPGFARGPS 257
 181 GPPGAGPGEGQGVGDLGAP-----GPGSPA-----GEPGPT 213
 258 GPGGPGPPGPKGSGEGFCAPGKGTGAKGPPGVGVQPPGAGGEGKRGARGEPGT 317
 214 GLPGRGPGGPGSGRFGPGADGVAGPKGAGRGSPGAGPKGSPGAGRGCEAGLPGAK 273
 318 GLPGRGPGGPGSGRFGPGADGVAGPKGAGRGSPGAGPKGSPGAGRGCEAGLPGAK 377
 274 GLTSGSPGPGDKTGPAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 333
 378 GLTSGSPGPGDKTGPAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 437
 334 GVPGPAGVAGPKDGEAGAQGPPGAGPAGRGSGPGAGSGPGQGLPGPAGPPGEGAKP 393
 438 GVPGPAGVAGPKDGEAGAQGPPGAGPAGRGSGPGAGSGPGQGLPGPAGPPGEGAKP 497
 394 GEQGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
 498 GEQGVPGDLGAPGSPGAGERGFGPBERGVQPPGAGPRGANGAPGNDGAKGDAGAPAP 557
 424 GERGPGSRGFGADGVAGPKGAGRGSPGAGPKGSPG-----AGRPGEGALP 474
 558 GSQGAPGLQGMPCERGAALPGPKDGRDAGAPKGDGSPGKDGVLGTGTPGPPGAP 617
 475 GAK-----GLTSGSPGPGDKTGPAGQDGRGPPGPPGARGQ----- 515
 618 GPKGSGSPGAGPTGARGAPGDRGEPGPPGAGPAGPPGADGQPGKAGGPDGAKGDA 677
 516 -----AGVMGPPGPKGAAGEPKAGRGVPGPPGA-----VGPAGKDGAEAGQPP 561

678 GPPGAGPAGPPGPIGNVAPGAPKARGSGAGPCATGTPGGAAGRVGPPPSGNAGPPGP 737
 562 GPAGP---AGERGQAGSPGFGQLGPGAGPPEAGKPGEQGVPGDLGAPPSGPAGEP 618
 738 GPAGKEGGKPRGTPGAPRGPEVGPFPGPPGAPGKESPGADGPAGAPGTGPGQIAGQR 797
 619 GPTGLPGRGPGERGGSGRFGADGVAGPKGAGRGSPGAPG---KGSPEAGRPCEA 675
 798 GVVGLPQGRGPGFGLPGSGERGKQKQPSGASGERGPPGPMGLAGPFGESGREGAP 857
 676 GLPKAGLTVSGSPGPDGKTKTPGPGAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGEP 735
 858 GAEGSPCRDGSAGKGRGTGTPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 917
 736 GKAGRGVPGPPGAVGAPKAGE-----AGAQQGPPG-----AGPA 771
 918 GPAGAGPAGPQGRGDKGTGEGQDGRGKGRHFGSLGQGGPPGPPSGPGEQPGSGAGPA 977
 772 GERGQAGPAGSP---GFGGLPGPAGPPEAGKPGEQGVPGDLGAPGSPGAPAG 820
 978 GPRPPGSAAPGKDGKGLNGLPGIPGPRGRTGDAGPVGPPGPPGPPGPPG 1029

RESULT 8
 AAY84538
 ID AAY84538 standard; protein; 1171 AA.
 XX
 AC AAY84538;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
 XX
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW ss. transforming growth factor-beta1; TGF-beta1; chimera.
 XX
 OS Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 858 /note= "Gly encoded by GCT"
 FT
 XX
 FN EP992586-A2.
 XX
 PD 12-APR-2000.
 XX
 PP 07-OCT-1999; 99EP-00119184.
 XX
 PR 09-OCT-1998; 98US-00169768.
 XX
 PA (USU) US SURGICAL CORP.
 XX
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 PI N-PSDB; AAA12498.
 DR WPI; 2000-259138/23.
 XX
 DR N-PSDB; AAA12498.
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 PS Claim 23; Fig 15; 260pp; English.
 XX
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC produce extracellular matrix proteins. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell,
 CC incorporating the nucleic acid sequence into the cell, and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention

incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents chimeric collagen 1 (alpha1)/transforming growth factor-beta1 (TGF-beta1) protein, which may be produced using the method of the invention

XX SQ Sequence 1171 AA;

Query Match		59.6%; Score 2765; DB 3; Length 1171;
Best Local Similarity		55.4%; Pred. No. 1.3e-162;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;		
QY	1	GPFGPGTGLPFPGERGSGRFGPGADGVAGPKGAGRGSPGAPGKSGPGEA--- 57
DB	18	GPWGSPGRLGPGPGAGPGQGPGGEGFEGASGPMGPRPFPKNGDDGAGKP 77
QY	58	GRFGAELP---GAKGLTSGSPGP-----DKTTPPGAGDGRPPGPPGAR 105
DB	78	GRFGGPPPGQAGRLGPGTAGLFGMKHGRFSGLDCAKADAGPAGPKGPGSPGENGAP 137
QY	106	GQAGVMGFPKGAAGSPGKAGR-----GVPPGNAVGA-----GKDGAQAQ 150
DB	138	GQMGRLPGERGPGAPGARGNDGATGAAGPPGPTGPPAGPFPFAVAKGASGPQ 197
QY	151	GPFGPAGPAGERGEQGP-----AGSPGFQGLGPA 180
DB	198	GPRGSEGPQVGRGPPGPAAGPAGNPDGQPGKAGKANGAPGAGPFGARGPS 257
QY	181	GPFGAELPGEQVFDLGNP-----GPSGA-----GEPGPT 213
DB	258	GPQGGPGPGPKGNSGEPGAGSKGTGAKGEPGVGVQPPGPAEGEGKARGEPGPT 317
QY	214	GLPFPGERGSGRFGPGADGVAGPKGAGRGSPGAPGKSGPGEAGRGAGLPGAK 273
DB	318	GLPFPGERGSGRFGPGADGVAGPKGAGRGSPGAPGKSGPGEAGRGAGLPGAK 377
QY	274	GLTSGSPGPDGKTGPPGAGDGRPPPGARGQAGVMGFPKGAAGEPGKAGER 333
DB	378	GLTSGSPGPDGKTGPPGAGDGRPPPGARGQAGVMGFPKGAAGEPGKAGER 437
QY	334	GVPPPGAVGPAKDGGAAGQPPGAPGAGRGEGGPGAGSGFGQGLPGAPPGAGKP 393
DB	438	GVPPPGAVGPAKDGGAAGQPPGAPGAGRGEGGPGAGSGFGQGLPGAPPGAGKP 497
QY	394	GEQVFGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
DB	498	GEQVFGDLGAPGSPGAGERGPPGPGVQPPGAPRGANGANGKADGAGAP 557
QY	424	GRGGPGSRGFGADGVAGPKGAGRGSPGAPGKSGPGE-----AGRPGEAGLP 474
DB	558	GSQAGPGLQMPGERGAAGLPKGDGADGPKGADGSGDKGVRLTGTPTGPPGAPAG 617
QY	475	GAK-----GLTSGSPGPDGKTGPPGAGDGRPPGPPGARGQ----- 515
DB	618	GDKGSGSPGAPGTGARGAGDGRGPPGPPGAGFPAGPPGADGQPGKAGGPDAGAKDA 677
QY	516	-----AGVMGFPCKGAAGEPGKAGERGVGPPGGA-----VCPAGKDGGAAGQPP 561
DB	678	GPPGAPGAPGPPGPIGNVCAKAGKARGSAGPPGATGFPGAAGRVGPPGSGNAGPPFP 737
QY	562	GPAGP-----AGEGEGPAGSPFQGLPGAPGPPGAGKPGEGQVGDGLGAPGSPGAPG 618
DB	738	GPAGKGGKPRGETGTPAGRPGEVPPGPPGAPGKSGPAGDGPAGAPGTFPGQIAGQR 797

QY	619	GPTGLPGERGSGRFGPGADGVAGPKGAGRGSPGAPG---KSGPGEAGRPGEA 675
DB	798	GVVGLPGERGSGRFGPGADGVAGPKGAGRGSPGAPG---KSGPGEAGRPGEA 857
QY	676	GLPQAGKGLTSGSPGPDGKTGPPGAGDGRPPPGARGQAGVMGFPKGAAGEBP 735
DB	858	GAEGSPGRDGSFKAAGDRGETGFPAGFPAGAPGAPGVPVGPAGKSGDRGETGFPAGPV 917
QY	736	GKAGRGVPPGPAVGPAGKDGSE-----AGAQQPPGP-----AGPA 771
DB	918	GPAGARGPAGQPGPRDGETGEGQDGRGKIHGRGFGSLQGPFPSPGSPGEGPSCASGPA 977
QY	772	GERGQGPAGSP---GFGQLPGPAGPCEAGKPGGQVPGDLAGPSPGSPAG 820
DB	978	GPRGPPGSAGAPGKDLNGLPPIGPPGPRGRTGDAGVPFPFPFPFP 1029

RESULT 9
AAW68485

ID AAW68485 standard; protein; 1464 AA.

XX AC AAW68485;

XX DT 08-DEC-1998 (first entry)

XX XX Human recombinant collagen protein.

XX KW Primer; PCR; amplification; human; collagen; mammal; plant; prosthesis;
KW cardiac valve; ligament; tendon; skin; gingival implant; perfumes;
KW nerve regeneration; antibiotic; growth factor; cancer; inflammatory;
KW gelatin; glue; food.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..22

FT Protein /note= "signal peptide"

FT Protein /note= "mature protein"

FT Cleavage-site 161

FT Cleavage-site /note= "cleavage site for aminopeptidase"

FT Cleavage-site 1218

FT Cleavage-site /note= "cleavage site for carboxypeptidase"

XX W09827202-AL.

XX PD 25-JUN-1998.

XX PF 17-DEC-1997; 97WO-FR002331.

XX PR 17-DEC-1996; 96FR-00016224.

XX PA (BIOC-) BIOCEM SA.

XX PI Gruber V, Exposito J, Ruggiero F, Conte J, Garrone R, Merot B;

XX PI Bournat P;

XX DR WPI: 1998-362771/31.

XX DR N-PSDE; AAV60814.

XX XX New recombinant nucleic acid for expressing collagen or derivatives in plants - useful as, e.g. bio-materials and in therapeutic, cosmetic and odontological compositions.

XX PS Disclosure; Fig 7; 138pp; French.

XX CC The invention relates to the production of mammalian collagen in plants. 2 clones: alpha3 and alpha22, spanning the human collagen type I gene were isolated from a MG-63 osteosarcoma library. Clone alpha3 contained 83 bp of the 5' untranslated region and the first 1920 bp of coding sequence, whereas clone alpha22 contained sequence encoding amino acids

Query Match	59.6%	Score	2765	DB 2	Length	1464						
Best Local Similarity	55.4%	Pred. No.	1.6e-162									
Matches	561	Conservative	31	Mismatches	228	Indels	192	Gaps	19			
Qy	1	GPPGEPGPTGLPGPPGRRGGPGSRGFP	GADGVAGPKG	PAGERGSPG	PAGPKGSPG	GEA---	57					
Db	179	GPMGPSGRGLGPPGAGPDQPGFQGGPP	GEPGEGASG	PMWPRGPPGPPG	CKNGDDGE	AGKP	238					
Qy	58	GRPEAGLP----	GAKGLTGS	PGSGPP	-----	DGKTGPPG	PAGQGRPP	PGPPG	PGAR	105		
Db	239	GRPGERGPPGQAGRLPGTLAGLPMK	GHRRGSLG	DGAKDGAP	GAPGKPE	SPSPENG	AP	298				
Qy	106	GQAGVMGPPGKAAGEPKGAGER----	GVPGPPG	AVGPA-----	CKDGE	AGAQ	150					
Db	299	GQMGRLGPERGRPGA	GPAGARG	DGATG	NAAGPPG	TGPAGPGF	CAVCAK	GEAGPQ	358			
Qy	151	GPPGPAGPAGRGEGPP-----	-----	AGSPQFQGL	PGPA	180						
Db	359	GPRSEGGVGRGEPGPPGAGAGPAG	NPAGDQF	GAKGANG	APCIAG	APFP	GARGPS	418				
Qy	181	GPPGEAGKPGSQGVPGDLGAP----	-----	GPSGA-----	-----	GEPGP	213					
Db	419	GPOGPGPPGPKNSGEFGA	PGSKDTC	AKGEPG	PVGVQGP	PPCPAGEB	KRGAR	EGPT	478			
Qy	214	GLPGPPGRRGPGSRGFP	PGADGVAG	PKGPAG	SGSPG	PAGPKGSP	GAEAGL	PGAK	273			
Db	479	GLPGPPGRRGPGSRGFP	PCADGVAG	PKGPAG	SGSPG	PAGPKGSP	GAEAGL	PGAK	538			
Qy	274	GLTSGPSGPDGKTGTPG	PPAGQDGR	PGPPG	PARGQAG	VMGFPG	KGAAGE	PKGAGER	333			
Db	539	GLTSGPSGPDGKTGTPG	PPAGQDGR	PGPPG	PARGQAG	VMGFPG	KGAAGE	PKGAGER	598			
Qy	334	GVPGPAGVGPAGKDG	EAGAAQ	PGPPG	PAGRG	EQGPAGS	PGFQGL	PGPAGPP	GEAGKP	393		
Db	599	GVPGPAGVGPAGKDG	EAGAAQ	PGPPG	PAGRG	EQGPAGS	PGFQGL	PGPAGPP	GEAGKP	658		
Qy	394	GEQGVPGDLGAPG	SPSGPAGE	-----	-----	PGP-----	-----	TGL	PGPP	423		
Db	659	GEQGVPGDLGAPG	SPSGAGER	FPGER	GVQGP	PPGAPR	GANGA	PGND	AKGD	AGAP	718	
Qy	424	GERGGPGSRGFP	PGADGVAG	PKGPAG	ERSG	PGPAG	KGSPGE	-----	AGR	PG	EGLP	474

CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;
 Best Local Similarity 55.4%; Pred. No. 1.6e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;
 QY 1 GPPGPGTGLPGRGSGRGGFPGADGAGVAGPKGAGRGSPGAGPKGSGPGEA--- 57
 DB 179 GPMGSGRGLPGLPGAGPQGFQPGPEGPGAGPMGPRGPPGKNGDDGEACKP 238
 QY 58 GRPGEAGLP---GAKLGTSGSPGSP---DGKTGPPGAGQDGRGPPGPPGAR 105
 DB 239 GRPGRGPPGQAGRLPGLTAGLPGMKHGRGSLDGAKGAGPAGPKGPGSPGENGAP 298
 QY 106 GOAGVMGPPGKAGAGEPKKAGER-----GVPPGPGAVGA-----GKGEAGAQ 150
 DB 299 GQMPRGLPGRGRPGAPGARGNDGATGAAGPPPTGAGPPGPGVAGKAGEAGPQ 358
 QY 151 GPPGPGAGPAGERGEGP-----AGSPGFGQLPGPA 180
 DB 359 GPRGSGPQVGRGPPGAGAGPAGNPGADGQPGAKGANGAPGIAGAPGFFGARGPS 418
 QY 181 GPPBAGKPGGQVPGDILGAP-----GPSGPA-----GEPGPT 213
 DB 419 GPQGRGPPGPKNSGSEFGAPGSGDGTCAKGPBGVGVQPPGAGRGKRGARGEPGPT 478
 QY 214 GLPGRGPPGSGRGGFCAGDVAGPKGAGRGSPGAGPKGSGPGEAGRLPGAK 273
 DB 479 GLPGRGPPGSGRGGFCAGDVAGPKGAGRGSPGAGPKGSGPGEAGRLPGAK 538
 QY 274 GLTSGPSGPPDGKTPGPPAGQDGRGPPGARGQGVGMGPPGKGAAGEPGKAGER 333
 DB 539 GLTSGPSGPPDGKTPGPPAGQDGRGPPGARGQGVGMGPPGKGAAGEPGKAGER 598
 QY 334 GVPGPAGVAGPAGKDGAGAGQPPGAGPAGRGSPGAGSPGFGPLFPAGPPGAGCKP 393
 DB 599 GVPGPAGVAGPAGKDGAGAGQPPGAGPAGRGSPGAGSPGFGPLFPAGPPGAGCKP 658
 QY 394 GEOGVPGDLGAPGSPGAGE-----PGP-----TCLPGPP 423
 DB 659 GEOGVPGDLGAPGSPGAGEGFGPGRGVQPPGPPGANGAPGNDGAKGDAGAPAG 718
 QY 424 GERGGPGRGPPGADGAVAGPKGAGRGSPGAGPKGSPGE-----AGRPEAGLP 474
 DB 719 GSQAGPGLQGMPEGAGLPGKDGADGAPKGDGSPGKDGVLGTGTPGPPGAPAG 778
 QY 475 GAK-----GLTSGPSGPPDGKTPGPPAGQDGRGPPGPPGARGQ----- 515
 DB 779 GPKGSGSPGAGTPGARGAGPDRGPPGPPGAGPPGAGPPGADGQFGKAGEPCDAKGA 838
 QY 516 -----AGVMGPPGPKAGAGEPKKAGERGVGPPGPA-----VGPAGKGEAGAQGPP 561

DB 839 GPPGPGAGPDPPIGNVGAAPGAKGARGSGAGPPGATGPPGAGRVGPPSGNAGPPGPP 898
 QY 562 GPAGP-----AGERGQGPAGSPGFGPLGPPAGPGEAGKPGEGQGVGDLGAPGSPGAGEP 618
 DB 899 GPACGKGGKPRGTGTGPRGPGVGVGPPGPPGAGKSGPAGADGAPAGTGTGPGQGIAGQR 958
 QY 619 GPTGLPGRGPPGGRGPPGSGRGGFPGADGAVAGPKGAGRGSPGAGP---KGSPPGAGRPGEA 675
 DB 959 GVVGLPQGRGPPGPPGPPGSGRGGKQKQPSGASGERGPPGPMGPPGLAGPPGSGREGAP 1018
 QY 676 GLPAGKGLTSGPSGPPDGKTPGPPAGQDGRGPPGPPGARGQGVGMGPPGPKGAAGEP 735
 DB 1019 GAEGSPGRDSSPGAKGDRGTGTGPPGAGPAGPAGPAGPAGPAGKSGDRGTGTGPPAGPVP 1078
 QY 736 KXAGRGVPPGPPGAVGPPAGKXGGE-----AGAQQGPPG-----AGPA 771
 DB 1079 GPAGARGPAGPQGRGDKGTGEGQDGRGKGRGFGSLQGGPPGSPGEGQSPGASGPA 1138
 QY 772 GERCEQGPAGSP---GFQGLPGRGPPGAGPGEAGKPGEGQGVGDLGAPGSPGAPAG 820
 DB 1139 GPRGPPGAGAGKXGDKLGLPPIGPPGPRGRTGDAGVPPGPPGPPGPPG 1190

RESULT 12
 ADD45051
 ID ADD45051 standard; protein; 1464 AA.
 XX
 AC ADD45051;
 DT 29-JAN-2004 (first entry)
 XX
 DE Human Protein P02452, SEQ ID NO 10483.
 KW Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX
 OS Homo sapiens.
 XX
 FN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002MO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GSHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 XX WPI; 2003-268312/26.
 DR GENBANK; P02452.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 CC
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially

expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at <http://wipo.int/pub/published.pct> sequences.

Sequence 1464 AA;

Query Match 59.6%; Score 2755; DB 7; Length 1464;
Best Local Similarity 55.4%; Pred. No. 1.6e-162;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY	1	GPPEPPTGLPGPPBGRGSGRFGADGVAGKPGAPGERGSPGAPKGSPEA---	57
Db	179	GPMPSPGRGLPGPPGAPGPGQFCPPCEPGEPCASGPMGRPGPGPKNGDDGEAGK	238
QY	58	GRPEAEAGL--GAKGLTGSFGSGP-----DGKTGPPAGODGRGPPGPPGAR	105
Db	239	GRGERPPPGQANGULPGTAGLPGMKHGRFGSLDGAKGDAGPAGPKGPGSGENGAP	298
QY	106	GQAGVMGFPKPKGAAGEPKGAGER-----GVFPFPGAVGPA-----GKDGEAGAQ	150
Db	299	GQMGPRGLPGERGRPGAPGAPAGARGNDGATGAAGFPPTGPAGPPGPPGAVGAKGEAGPQ	358
QY	151	GPPEPAPAGERGEOGP-----AGSPFQGLPGPA	180
Db	359	GPKGSEPGVGRGEPGPPGAGAGPAGNPGADGQPGAKGANGAPGTAGAPGPPGARGPS	418
QY	181	GPPEACKPEQGVPGDLGAP-----GFSGPA-----GBPGPT	213
Db	419	GPQPGPPGPKNSGEPGAPGSKGDTCAKEGPPGVQPPGPAGEGKRGARGERPPT	478
QY	214	GLPGPPGERGGPSRPPPGADGVAGPKGPAGERGSPGAPKGPSPEAGRPGGAGLPGAK	273
Db	479	GLPGPPGERGGPSRPPGADGVAGPKGPAGERGSPGAPKGPSPEAGRPGGAGLPGAK	538
QY	274	GLTGPSGSPDDGKTGPPGAGODGRPGPPGPCARGQAGVMGFPCKGAAGRPGKAGER	333
Db	539	GLTGPSGSPDDGKTGPPGAGODGRPGPPGPCARGQAGVMGFPCKGAAGRPGKAGER	598
QY	334	GVFPFPGAVGPAKGDGEAGAQGPPGPAGPAGERGEOGFAGSPGFQGLPGPAGPFPBGAKP	393
Db	599	GVFPFPGAVGPAKGDGEAGAQGPPGPAGPAGERGEOGFAGSPGFQGLPGPAGPFPBGAKP	658
QY	394	GEOGVPGDLGAPGSPGAGE-----GPP-----TGLPGPP	423
Db	659	GEQVPGDLGAPGSPGARGRPFGRGVQPPGPAGFRGANGAPGNDGAKGDAGAPGAP	718
QY	424	GERGPGCSRFPFGADGVAGPKGPAGERGSPGAPGKSGPCE-----AGRPGGAGLP	474
Db	719	GSQCAPGLQMPGERGAAGLPGPKGRDGDAGPKGADGSPGKDGVRGLTGTLPFGPPGAGP	778
QY	475	GAK-----CLTSPGSPGPDGKTGPPGPAQODGRPGPPPGPARGO-----	515
Db	779	GDKGESGPGAPGPTGARGAPGDRGPPGPPGAFAGPPGADGQPGAKGSPGDAGAKGDA	838
QY	516	-----AGVMGFPKPKGAAGBPKKAGERGVPPGPA-----VCPAGKDGAGAGQPP	561
Db	839	GPDPGAPGAPPGPIGNVGAAGKARGSGNPPGATGTFPGAARGVPGPPGSGWAGPPPPP	898
QY	562	GPAGP---AGERGEOQPAGSPGFQGLPGPAGPPGEGAGKPGEOGVPGDGLGAPGSPGAPGR	618

Db	399	GPAGEKGGKPRGRTGTGAPRGEVGGPPGPPAGEKSGPADGPAGAPGTPGPGQGIAGQR	958
QY	619	GPTGLPDPGERGCGPSGRGFPFADGVAKPGKAPAGERGSPGAP-----KSGSPAGRRPGEA	675
Db	959	GVVGLPQQRGERGFPGLPGPSGPGKQGPSASGERGPPGPMGPPGLAGPPGSGRGAP	1018
QY	676	GLPCAAGLITGSPSGPDPGKTPGPPGAPQDGRGPPGPPGARGAQVGMGPPGPKGAAGEP	735
Db	1019	GAEGSPGRDSCPAKGRDGTGTGAPGPGAPGAFCAFGVPGAGKSGDRGTGTGAPGAPGV	1078
QY	736	GKAGERGVPPGPGAVGAPGADGE-----AGAQQPPGP-----AGPA	771
Db	1079	GPAGARGAPGQPRGDKGTGTGQEDGRIKCHRGFSLGLOQPPGPPGSGQPSGASGPA	1138
QY	772	GERGEGQAGSP---GFQGLPGPAGPPGEAGKPGEGVPGDLAGAPGSPGAG	820
Db	1139	GPRGPPGSGAGAPKDGKGLNGLPSTGTPGPPRGRTGDAGPVGPPGPPGPPGPPG	1190

RESULT 13
ADR16800

ADRI6800
ID ADR16800 standard; protein; 1464 AA.

AC ADR16800;

DT 21-OCT-2004 (first entry)

XX
DE Human collagen I alpha1 (I) chain protein.

XX
KW
T-lymphocyte; AAA; abdominal aortic aneurysm; human; collagen I; therapy.

AA Homo sapiens.

XX
PN
US2004151731-A1.

XX
PD 05-AUG-2004.XX
PF 04-FEB-2003:XX
PR 04-FEB-2003: 2003UIS-00357851XX
PA
(JTCCH / JTCCH D IXX
 BT
 Ticks Dr.XX
DB WDT: 2004-EC147A/5A

XX
 DE
 Now, I closed my eyes, and

tissues useful to prevent/treat an abdominal aortic aneurysm or rupture.

PS Disclosure; SEQ ID NO 1; 24pp; English.

The present invention relates to a T-lymphocyte derived from abdominal aortic tissue which are specifically reactive with collagen I, collagen III or their fragments. The invention is useful to prevent or treat an abdominal aortic aneurysm (AAA) or rupture in a mammal. The present sequence is human collagen I alpha (I) chain protein. This sequence is used in the invention.

Sequence 1464 AA;

Query Match	59.6%;	Score 2765;	DB 8;	Length 1464;
Best Local Similarity	55.4%;	Pred. No. 1.6e-162;		
Matches 561; Conservative	31;	Mismatches 238;	Indels 192;	Gaps 19;

QY 1 GPPGPGPTGLPGPPGERGGPSRGFPCADGVAGPKGPAGERGSPGPAGPKGSPGEA --- 57

Db
179 GPMGPSGRGLPGPPGAPGPGQFGPPGEPGEFGASGPMGPRGPPGPPGKNGDDGEAGKP 238

QY 58 GRPGEAGLP---GAKGLTGSPGSPGP-----DGKT'GPPGPGAGODGRPGPPGPPGAR 105

239 GRPGERGPPGPOCARGLPCTAGLPGMKGHRGFSGLDGAKGADGAPGKGPSPGENGAP 298

106 GOAGVMGPPGPKGAAGEGKAGER-----GVPPGPGAVGPA-----GKDGEAGQ 150
 Db 299 GQMGPRGLGPRGRGAPGAPAGARGNDGATGAAGPGTGPAGPPGPGAVGAKGAGPQ 358
 Qy 151 GPPGPGAGPAGERGEGP-----AGSPFGQGLPGPA 180
 Db 359 GPRGSEGVGVRGPPGPPGAPAGAGPAGNPGADGQPGAKGANGAPGAGPAGPFGARGPS 418
 Qy 181 GPPGAGKPGGQVPGDLGAP-----GPSGPA-----GERGPT 213
 Db 419 GPQPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPGVQGGPPGAGEGKRGAGEPPT 478
 Qy 214 GLPDPGPRGRGPPGPPGADGAVGPKGAGERSGPGAGPKGSPGAGRPGAGLPGAK 273
 Db 479 GLPDPGPRGRGPPGPPGADGAVGPKGAGERSGPGAGPKGSPGAGRPGAGLPGAK 538
 Qy 274 GLTSGSPGPDGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 333
 Db 539 GLTSGSPGPDGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 598
 Qy 334 GVPDPGAVGAGKDGEGAGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 393
 Db 599 GVPDPGAVGAGKDGEGAGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 658
 Qy 394 GEOGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
 Db 659 GEOGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
 Qy 424 GERGPPSRGPPGADGAVGPKGAGERSGPGAGPKGSPGAGRPGAGLPGAK 474
 Db 719 GSQAPGLQMGPPGREGAAGLPGPKDGRDAGPKGADSGKDGVRGLTGPFGPPAGAP 778
 Qy 475 GAK-----GLTSGSPGPDGKTGTPGPPGAGQDGRPPGPPGARGQ-----515
 Db 779 GDKGESGPPGAPGTGARGAPGDRGEPGPPGPPGAGPPGADGQPGAKGEPGDAGAKGDA 838
 Qy 516 -----AGVMGPPGPKGAAGEGKAGERGVPPGPA-----VGPAGKDGAGAGGPP 561
 Db 839 GPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 898
 Qy 562 GPAGP-----AGRGEQPGAGSPGQGLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 618
 Db 899 GPAGKEGKGRGTGTPAGRGEVGPFGPPGPPGAGEKGSFGADGAGPAGTGTGPGIAGQ 958
 Qy 619 GPTGLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 675
 Db 959 GWVGLPGQGRGERGPPGLPGSPGPKQPGSPGASGERGPPGPPGPPGPPGPPGPPGPP 1018
 Qy 676 GLPGAKGLTSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 735
 Db 1019 GAEGSPGRDGSFGAKGDRGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1078
 Qy 736 GKAGERGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 771
 Db 1079 GPAGARGPAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1138
 Qy 772 GERGEGPAGSP-----GFGGLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 820
 Db 1139 GPRGPPGAGAPGKDGGLNGLPGIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1190

RESULT 14

ADRI16425
 ID ADRI16425 standard; protein; 1464 AA.
 XX AC ADRI16425;
 XX XX
 DT 21-OCT-2004 (first entry)
 XX DE
 XX DE Human collagen I alpha (I) chain protein.
 KW T-lymphocyte; AAA; abdominal aortic aneurysm; vaccine; therapy; human; collagen I.

XX Homo sapiens.
 OS US2004151732-A1.
 FN 05-AUG-2004.
 PD 04-FEB-2003; 2003US-00358024.
 PF 04-FEB-2003; 2003US-00358024.
 PR (JICH/) JICHA D L.
 PA (PELU/) PELUSE S.
 XX Jicha DL, Peluse S;
 DI WPI; 2004-570707/55.
 DR Isolated or purified lymphocytes derived from blood in abdominal aortic
 XX aneurysm patients useful for early diagnosis of aneurysms are reactive
 PT with collagen I, collagen III and/or their fragments.
 PT Disclosure; SEQ ID NO 1; 22pp; English.
 PS The present provides a T-lymphocyte derived from blood in abdominal
 CC aortic aneurysm (AAA) patients which are reactive with collagen I,
 CC collagen III and their fragments. The invention is useful in diagnosis,
 CC prevention and treatment of abdominal aortic aneurysm or rupture in a
 CC mammal. The invention is also useful in the vaccine preparation. The
 CC present sequence is human collagen I alpha (I) chain protein.
 XX Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 8; Length 1464;
 Best Local Similarity 55.4%; Pred. No. 1.6e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;
 Qy 1 GPPGEPPTGLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 57
 Db 179 GPMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 238
 Qy 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTGPPGAGQDGRPPGPPGPPGPPGPP 105
 Db 239 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTGPPGAGQDGRPPGPPGPPGPPGPP 298
 Qy 106 GQAGVMGPPGPKGAAGEGKAGER-----GVPPGPGAVGPA-----GKDGEAGQ 150
 Db 299 GQMGPRGLGPRGRGAPGAPAGARGNDGATGAAGPGTGPAGPPGPPGAVGAKGAGPQ 358
 Qy 151 GPPGAPGAGERGEGP-----AGSPFGQGLPGPA 180
 Db 359 GPRGSEGVGVRGPPGPPGAPAGAGPAGNPGADGQPGAKGANGAPGAGPAGPFGARGPS 418
 Qy 181 GPPGAGKPGGQVPGDLGAP-----GPSGPA-----GERGPT 213
 Db 419 GPQPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPGVQGGPPGAGEGKRGAGEPPT 478
 Qy 214 GLPDPGPRGRGPPGPPGADGAVGPKGAGERSGPGAGPKGSPGAGRPGAGLPGAK 273
 Db 479 GLPDPGPRGRGPPGPPGADGAVGPKGAGERSGPGAGPKGSPGAGRPGAGLPGAK 538
 Qy 274 GLTSGSPGPDGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 333
 Db 539 GLTSGSPGPDGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 598
 Qy 334 GVPDPGAVGAGKDGEGAGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 393
 Db 599 GVPDPGAVGAGKDGEGAGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGPKAGER 658
 Qy 394 GEOGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
 Db 659 GEOGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423

Search completed: June 17, 2005, 15:08:26
Job time : 140.148 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	2765	59.6	1464	4	US-09-331-347C-21	Sequence 21, Appl
2	2752	59.3	1461	4	US-09-585-887-9	Sequence 9, Appl
3	2752	59.3	1461	4	US-09-289-578-9	Sequence 9, Appl
4	2746	59.2	1057	3	US-08-931-820-1	Sequence 1, Appl
5	2704	58.3	1341	3	US-08-963-825-18	Sequence 18, Appl
6	2704	58.3	1341	3	US-09-500-811-18	Sequence 18, Appl
7	2704	58.3	1341	3	US-09-570-573-18	Sequence 18, Appl
8	2704	58.3	1341	3	US-09-548-608-18	Sequence 18, Appl
9	2511	54.1	822	3	US-09-219-849-49	Sequence 49, Appl
10	2488	53.6	1017	4	US-08-468-998-10	Sequence 10, Appl
11	2478	53.4	1064	1	US-08-642-253-62	Sequence 62, Appl
12	2476	53.4	1060	3	US-08-931-820-3	Sequence 3, Appl
13	2476	53.4	1418	3	US-08-963-825-20	Sequence 20, Appl
14	2476	53.4	1418	3	US-09-010-999-1	Sequence 1, Appl
15	2476	53.4	1418	3	US-09-500-811-20	Sequence 20, Appl
16	2476	53.4	1418	3	US-09-570-573-20	Sequence 20, Appl
17	2476	53.4	1418	3	US-09-548-608-20	Sequence 20, Appl
18	2464	53.1	1442	2	US-08-316-650-12	Sequence 12, Appl
19	2464	53.1	1442	5	PCT-US95-02251-12	Sequence 12, Appl
20	2370	51.1	1057	3	US-08-931-820-4	Sequence 4, Appl
21	2350.5	50.7	1078	3	US-08-963-825-21	Sequence 21, Appl
22	2350.5	50.7	1078	3	US-09-500-811-21	Sequence 21, Appl
23	2350.5	50.7	1078	3	US-09-570-573-21	Sequence 21, Appl
24	2350.5	50.7	1078	3	US-08-948-608-21	Sequence 21, Appl
25	2267.5	48.9	1366	3	US-08-963-825-19	Sequence 19, Appl
26	2267.5	48.9	1366	3	US-09-500-811-19	Sequence 19, Appl
27	2267.5	48.9	1366	3	US-09-570-573-19	Sequence 19, Appl

; Sequence 9, Application US/09289578

; Patent No. 6428978

; GENERAL INFORMATION:

; APPLICANT: Olsen, David R

; APPLICANT: Chang, Robert

; APPLICANT: McMullin, Hugh

; APPLICANT: Hitzeman, Ronald A.

; APPLICANT: Chisholm, George

; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND

; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT

; TITLE OF INVENTION: CELLS

; FILE REFERENCE: 2250030400

; CURRENT APPLICATION NUMBER: US/09/289,578

; CURRENT FILING DATE: 1999-04-10

; PRIOR APPLICATION NUMBER: 60/084,828

; PRIOR FILING DATE: 1998-05-08

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1461

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-289-578-9

Query Match 59.3%; Score 2752; DB 4; Length 1461;

Best Local Similarity 55.2%; Pred. No. 2.2e-170;

Matches 559; Conservative 31; Mismatches 230; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGPPGRRGGGSGRGGFPGADGVAGPKGAGRGSPGAPGKSGPGEA--- 57

DB 176 GPMGSGRGLPGPPGAPGQFQFPPEPGEAGSGPMGPRGPPGPKNGDDGEAKP 235

QY 58 GRPGGAGLP---GAKGLTSGPSGPGP-----DGKTGPPGAGQDRPFPGPAGAR 105

DB 236 GRPGGPPGPGQAGRLPGTAGLPCMKHGRGSLDGAKGADGAPGAPGKGEPSGPGCAP 295

QY 106 GQAGVMGPPGKGAAGEPGKAGER-----GVPGPGAVGA-----GKDGGAQA 150

DB 296 GQMGPRGLPGRRGPGAPGAGARGNDGATGAAGPPGPTGAPGPPGPGAVGAKGAGPQ 355

QY 151 GPPGAPGAGRGEGGP-----AGSPGFGQLPGPA 180

DB 356 GPRGSEGGVGRGPPGPPGAGAGPAGNPGADQFGAKGANGAPGIAGAPGPPGARGPS 415

QY 181 GPPGAGKPGEGVPGDLGAP-----GPSGPA-----GEPGPT 213

DB 416 GPGGPGPPGKNSGSEFGAPGSKGDTGAKGPPGVQGGPPGAPGEGKRGAREPPT 475

QY 214 GLPGPPGRRGGSGRGGFPGADGVAGPKGAPGAGRGSPGAPGKSGPGEAGRLPGAK 273

DB 476 GLPGPPGRRGGSGRGGFPGADGVAGPKGAPGAGRGSPGAPGKSGPGEAGRLPGAK 535

QY 274 GLTSGPSGPDGKTGPPGAPGQDRPDPGPPGARGQAGVMGPPGKGAAGEPGKAGER 333

DB 536 GLTSGPSGPDGKTGPPGAPGQDRPDPGPPGARGQAGVMGPPGKGAAGEPGKAGER 595

QY 334 GVPGPPGAVGAPGKDGGAQAQPPGAPGAPGAGRGQAGSGPFGQLPGPAGPPGEGAGKP 393

DB 596 GVPGPPGAVGAPGKDGGAQAQPPGAPGAGRGQAGSGPFGQLPGPAGPPGEGAGKP 655

QY 394 GQGVGPDGLGAPGPPGSPAGE-----PGP-----TGLPGPP 423

DB 656 GQGVGPDGLGAPGPPGSPAGE-----PGP-----TGLPGPP 715

QY 424 GERGGPSRGGFPGADGVAGPKGAPGAGRGSPGAPGKSGPGE-----AGRPGGAGLP 474

DB 716 GSQGAFLGQMGFGERGAAGLPFGKDRGDAGPKGADSGFGKDGVRGLTGPTGPPGAPAG 775

QY 475 GAK-----GLTSGPSGPDGKTGPPGAGQDRPDPGPPGARGQ----- 515

DB 776 GDKGSGSGPAGPTGARGAPGDRGEPGPPGAPGAGPPGADGQGKAGEPGDAGAKGDA 835

QY 516 -----AGVMGPPGKGAAGEPGKAGERGVGPPGPA-----VPGAKDGGAQAQPP 561

DB 836 GPPGAPGAPGPPGPIGNVGAFAKAGSGAGPPGATGFPAGAGRVGPPGSGNAGPPGPP 895

QY 562 GPAGP---AGRGGQGPAGSPGFGGLPGAPGPPGAGPGEAKPKGQGVGDLGAPGPPGAPGEP 618

DB 896 GPAGKEGKGRGRTGTPAGRPGEVGPFGPPGAPGAGEKSGPAGADGAPGAGTGTGPPQGIAGQR 955

QY 619 GPTGLRPPGRRGGSGRGGFPGADGVAGPKGAPGAGRGSPGAPG---KGSPEGAGRPGEA 675

DB 956 GVVGLPGQRRGRRGPPGLPGSGEPGKQKQPSGASGERGPPGPMGPPGLAGPPGSGREGAP 1015

QY 676 GLPGAKGLTSGPSGPDGKTGPPGAPGQDRPDPGPPGARGQAGVMGPPGPKGAAGEP 735

DB 1016 GAESGPRDGSFGAKGDRGETGTPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1075

QY 736 GKAGERGVGPPGAVGAPGAKDGE-----AGAQQGPPG-----AGPA 771

DB 1076 GPVGARGPAGPQGPGRDKGTGETGQDRIKIHGRGSLGQGGPPGPPGSGEGQSGASGPA 1135

QY 772 GERGEQGPAGSP---GFGQLPGPAGPGEAGKPGEGQGVFDLGAPGSGPAG 820

DB 1136 GPRGPPGSAGAPGKDGGLNGLPGTIGPFGPRGRTGDAGPVGPPGPPGPPGPPG 1187

RESULT 4

US-08-931-820-1

; Sequence 1, Application US/08931820

; Patent No. 6010863

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Assay for collagen degradation

; NUMBER OF SEQUENCES: 4

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/931,820

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 96202596.1

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1057 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; TISSUE TYPE: Collagen type I

US-08-931-820-1

Query Match 59.2%; Score 2746; DB 3; Length 1057;

Best Local Similarity 56.3%; Pred. No. 4.1e-170;

Matches 556; Conservative 32; Mismatches 232; Indels 168; Gaps 18;

QY 1 GPPGEP-----GPTGLPGPPGRRGGGSGRGGFPGADGVAGPKGAPGAGERGSP----- 45

DB 42 GPPGEPGPGASGPMGPRGPPGPGKNGDDGGEAKKPGRGPPGPPGQAGRLPGTAGLP 101

QY 46 -----GPAGPK---GSPGAGRPGGAGLPAGKGLTSGPSGPGP----- 80

DB 102 GMKGRHGRGSLDGAKGADGAPGAPGKGPSPGCPGAGPAGQMGPRGLPGERGRPGAPGAPACAR 161

QY 81 ---DKT---GPPGAPGQDRPDPGPPGAPGAGQA-----GVMGPPGPKGAAGEP 123

DB 162 GNDGATGAAGPPGPTGTPAGPFGPFGAVGAKGAGPQGRGSEGGVGRGEPGPPGAGAA 221

QY	124	GKAGRGVPCGAVG	PAGKDCGAGACQ	PPGPAGPAGERGEOG	PAGSPGFQGLPGPAGPP	183			
DB	222	GPAGNPGADQGP	GAKGANGAPGIAG	APGPGSPQGPCTG	AGLGVGPAGSK	281			
QY	184	GEAGKPGEOQV	PGDGLGAPPS	-----GPAGEPGT	GLPGPGRGGPGSGRGF	237			
DB	282	GESGNKEPGS	AGPQPGPSEEGK	RPNGEAGSAGPFG	PGLRGGPGSRGF	341			
QY	238	GPXGPAGERG	SPGAGKSGPGB	AGPGBAGLPGAK	GLTSGSPGPGDKTGP	297			
DB	342	GPXGPAGERG	SPGAGKSGPGB	AGPGBAGLPGAK	GLTSGSPGPGDKTGP	401			
QY	298	GRGPBGPPC	ARGOAGVMGFP	PGPKGAAGPGKAG	RGVPGPGAVGPAGK	357			
DB	402	GRFGPBG	PCARQAQVMG	FPGPKGAAGPGKAG	RGVPGPGAVGPAGK	461			
QY	358	GPAGPAGERG	EOGPAGSPQGL	PGPAGPGBAGK	PGEQVPGDLCAPPG	413			
DB	462	GPAGPAGERG	EOGPAGSPQGL	PGPAGPGBAGK	PGEQVPGDLCAPPG	521			
QY	414	-----PCP	-----	-----TGLPGP	PPGRGGPGSRGF	447			
DB	522	GERGVOPPG	PPGPRGANGA	PNDAKGDAGAPAG	PSQAGLQMPG	581			
QY	448	GERGSPGAP	KPGKSGPGE	-----AGRPG	EAGLPGAK-----GLTSGSPSGPD	489			
DB	582	GDRGDAGPK	GADGSPGK	DGVRGLTGP	PGPAGAPDGK	SGSPGAPGTCARGAPGDR	641		
QY	490	GKTGPPG	PAGQDGRPG	PPGPPGARGQ	-----	-----AGVMGFP	534		
DB	642	GERGPPG	PAGPAGPPG	ADQGPAGKGP	PDAGAKDAGP	PGPAGPAGPGPI	701		
QY	535	GERGVPPG	CA-----	-----VGPAGK	DGAGACQPPG	PAGP-----AGERGEOGPAGSPGFQ	582		
DB	702	GARGSGPP	CACTFP	PGAGRVGPPG	SGNAGPPGPGPAGK	SGKGRGETGTGAPRG	761		
QY	583	GLPGPAG	PPGGAACKP	GBEQVPGDLG	APGSPGAPBPGT	GLPGPGRGGPGSGRGFP	642		
DB	762	GPFGPGP	AGKEGSPADG	PAGAPGTPGQ	TAGORGVVGL	PGQRGERRGFP	821		
QY	643	GVAGPKG	PACERGSPG	PAGP-----	-----KGSFPG	BAGRPGEAGLPCAK	GLTSGSPGSPDGKCTGPP	699	
DB	822	GKOGPSG	SAGERPPG	PMGPPGLAGP	PGSGREGAPCA	EGSGFORPGSG	PKAKDRGETGPA	881	
QY	700	GPAGQDGR	PGPPPGAR	GQAQVMGPPG	KGAAGEPGKAG	RGVPGPGAVGPAGK	DGGE-	758	
DB	882	GPPCAP	PAGPAGP	VPVGPAGKSG	DRGETGTG	PAGPAGPVG	PAGARGPAGPQ	941	
QY	759	-----	-----ACACQ	PPG-----	-----	-----AGPAGERG	EOGPAGSP-----GFCGLPGPA	792	
DB	942	GDRGIKH	RGHFG	GLQGP	PPGPPGSG	PGQPSGASG	PAGPRGPPGSG	AGAPKQGLNGLP	1001
QY	793	GPFGA	EAKPGEOQ	VP	GDLAGP	SGSPGAG	820		
DB	1002	GPFGPRG	RTGDAG	VP	GP	PPGPPGPPG	1029		

RESULT 5

```

RESUMI 3
US-08-963-825-18
: Sequence 18, Application US/08963825
: Patent No. 6110689
: GENERAL INFORMATION:
: APPLICANT: Qvist, Per
: APPLICANT: Bonde, Martin
: TITLE OF INVENTION: A Method for Assaying Collagen Fragments
: TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
: TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
: TITLE OF INVENTION: Disorders Associated with the Metabolism of
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: darby & Darby PC
: STREET: 805 Third Avenue

```


Db	601	GLQGMPEGERGAAGLPGKGDGRDAGPKGADSPGKDGVRGUTGPIGPPGAPGPDGKES	660
Qy	478	-----GLTSGPSGPGDKTTPPPAGQDGRPPGPPGARGO-----A	516
Db	661	GPSGPAGPTGARGA PDRGEPGPPGAGFPAGADGQGAKEGPDGAGKGDAGPPGPA	720
Qy	517	GVWGPFPGKGAAGEPGKAGERGVPPGPA-----VGPAGKDGGAAGAGPPGAPG-	566
Db	721	GPAGPPGPTGNVAPCAKARGAGSAGPPGATGPPGAAGRVGPPGSGNAGPPGPPGPA	780
Qy	567	--AGERGEGPAGSGFOGLPGPAGPPGKGEQGVPGDLGAPGSPGAGEPGTCLP	624
Db	781	GKGGRGETGPRGPEVGPVPPGPPGAGEKSGPADGAPAGPTGPGIAGQGVWGLP	840
Qy	625	GPPEGRGPGSGRFGCADGVAGPKPAGERSGPPAGP---KGSPEAGRPGCEAGLPGAK	681
Db	841	GQGERGFFGLPGSGEPKQPSGASERGPPGPMGPPGLAGPPCESGREGAPGEGSP	900
Qy	682	GUTSGPSGPPDKTTPPGPAGQDGRPPGPPGARGQGVMGPPKGAAGEPGKAGER	741
Db	901	GRDGSFGAKGDRGETGPAGPPGACXAGAPGVGPAKSGDRGETGPAGPAGVPAGAR	960
Qy	742	GVPPPGCAVGPAGKGE-----AGAGPPG-----AGPAGERGEQ	777
Db	961	GPAGPQGRGDRGETGEOGDRGKIGHRGFSLQGPVPPGSGEQSGASGAPGRGPP	1020
Qy	778	GPAGSP--GFGGLPGPAGPPGCAKPGEGQGVPGDLGAPGSGGAP	820
Db	1021	GSAGAPKDGMLNGLPGPIGPPRGRTGDAGVPVPPGPPGPPG	1066

RESULT 6

US-09-500-811-18

Sequence 18, Application US/09500811

Patent No. 6323314

GENERAL INFORMATION:

APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin

TITLE OF INVENTION: A Method for Assaying Collagen Fragments

TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

TITLE OF INVENTION: Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/500,811

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: 08/187,319

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Gogoris, Adda C

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 4305/08701

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237

TELEX: 236687

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

QY 778 GPAGSP---GFQGLPGPAGPPGAGKPGEGQGVFGDLGAPGSPGAGP 820
 Db 1021 GSAGAPKCDLGLNGLPGIPGPRGRTGDAGVPGPFGPPGPPG 1066

RESULT 7

US-09-570-573-18

; Sequence 18, Application US/09570573
 ; Patent No. 6342361
 ; GENERAL INFORMATION:
 ; APPLICANT: Oviast, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/570,573
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1341 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (I)

US-09-570-573-18

Query Match 58.3%; Score 2704; DB 3; Length 1341;
 Best Local Similarity 57.3%; Pred. No. 2.6e-167;
 Matches 542; Conservative 28; Mismatches 244; Indels 132; Gaps 15;
 QY 1 GPRGE---PGPTGLPGPPGGRGSGRPF-----PGADGVAGPKGPAGER 42
 Db 127 GPPGRDIPGQGLPGPPGPPGPPGLGNGFNAPQLYGYDEKSTGGISVPGPMGSPGR 186
 QY 43 GSPGAPGPKGSGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGPGAGDGRP---GPP 99
 Db 187 GLPGPPGAPGZGFGPPGZGZPGASGPMGPRGPPGPKGBGZAKGPRGZRGPP 246
 QY 100 GPPGARGQGVNMGVFPKPGAAEPKAGERGVPVGPVAGVAGPKDGEAGAGPAGPAGPA 159
 Db 247 GPZGARGLPGTAGLPGMKHGRGFSGLBGAKGAPGAPKGPZGSPGZBGAPGZMGPFGPK 306

QY 160 GERGEQGPAGSPGFQGLPGPAGPPGCEAGKPGEQGVFGDLGAPGSPGAGP 219
 Db 307 GNSGE-----PGAPGSKGDTGAKGEPGVQGVQGPAGBEGKRGARCEPPTGLPGP 360
 QY 220 GERGGPSRGFPAGDGVAGPKGPAGERSPPGAGPKGSPGEAGRPGEAGLPAKAGLTGSP 279
 Db 361 GERGGPSRGFPAGDGVAGPKGPAGERSPPGAGPKGSPGEAGRPGEAGLPAKAGLTGSP 420
 QY 280 GSPGPDGKTGPPGAGODGRPPGPPGARGOAGVWMPGPKGAAGEPKGAGERGVP 339
 Db 421 GSPGPDGKTGPPGAGODGRPPGPPGARGOAGVWMPGPKGAAGEPKGAGERGVP 480
 QY 340 GAVGPAGKDGAGAGQPPGPPGAGPAGERGEQGPAGSPGFQGLPGPAGPCEAGKPGEQGV 399
 Db 481 GAVGPAGKDGAGAGQPPGPPGAGPAGERGEQGPAGSPGFQGLPGPAGPCEAGKPGEQGV 540
 QY 400 GDLNAPGSGPAGE-----PGP-----TGLPGPPGERGPP 429
 Db 541 GDLNAPGSGPAGERGKGFPPGPPGPPGARGGANGAPGNDGAKGDAGAPGAPGSGQAP 600
 QY 430 GSRGPPGADGVAGPKGPAGERSPPGAGPKGSPGE-----AGRPGEAGLPGAK--- 477
 Db 601 GLQGMFGERGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGPITGPPGAPGADKGES 660
 QY 478 -----GLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQ-----A 516
 Db 661 GPSGPAGPTGARGAPGDRGEPGPPGAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPA 720
 QY 517 GVMGPPGKGAAGBPKGAGRGVFPGPQA-----VGPAGKDGGEAGAGPPGPPGAP- 566
 Db 721 GPAGPPGPIGNVGAAPGAKGARGSPGATGFPGAAGRVGPPGSPGNAGPPGPPGAPGKE 780
 QY 567 --AGERGQGPAGSPGFQGLPGPAGPPGCEAGKPGEQGVFGDLGAPGSPGAGP 624
 Db 781 GKGPRGETGAPGEPGPPGPPGAGEKSPGADGAPGAPGTPGPGIAGQRVGLP 840
 QY 625 GPPGERGGPSRGFPAGDGVAGPKGPAGERSPPGAP----KGSFGEAGRPGEAGLPGAK 681
 Db 841 QQRGERGPPGLPGSPGEPKQKQPSGASGERGPPGPMGPPGLAGPPGESREGAPGAEESP 900
 QY 682 GLTSGPSGPDGKTGPPGAGODGRPPGPPGARGOAGVWMPGPKGAAGEPKGAGER 741
 Db 901 GRDSPGAKGDRGETGPPGAPGAXGAGPVPAGKSGDRGETGPPAGPAGVPVGPAGAR 960
 QY 742 GVPFPAGVAGPAGKDGE-----AGAQGPFGP-----AGPAGERGEQ 777
 Db 961 GPAGPQGRGDKGTGEGQDRGIKGRHGFSLQGPFPSPGQSGSAGSPAGPRGPP 1020
 QY 778 GPAGSP---GFQGLPGPAGPPGAGKPGEGQGVFGDLGAPGSPGAG 820
 Db 1021 GSAGAPKCDLGLNGLPGIPGPRGRTGDAGVPGPFGPPGPPGPPG 1066

RESULT 8

US-09-548-608-18
 ; Sequence 18, Application US/09548608
 ; Patent No. 6355442
 ; GENERAL INFORMATION:
 ; APPLICANT: Oviast, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:

Query Match	54.1%	Score 2511;	DB 3;	Length 822;
Best Local Similarity	56.6%;	Pred. No. 4.9e-155;		
Matches 498; Conservative	35;	Mismatches 191;	Indels 156;	Gaps 15;

Qy	1	GPPEGPTGLPPGE-----	RGPGSRGFGPADGVAGPKPAGERGSPGA	48
Db	12	GPAPAPGQFGPPGEPGEGSGPMGRPPGPPQKNGDDGEAKPGRGRRGPPGPQ	71	
Qy	49	GPKSGPGEARP-----	GEAGLPKAGKLTGSPGSPGDGKTGPPGPAGQD	93
Db	72	GARGLPCTAGLPCKMGRHSGFLDGAKGDAGPAGPKGEGSPGECNAGAPGQMGRGLPGER	131	
Qy	94	GRGPPGPPGARQAGVMGPPGPKGAAGE-----	PKAGERGVPPGPGVAGPAGKDGEA	147
Db	132	GRGPPPTAGRNGDCAVGAAGPGPTGTPGPFPGVAGKAGEAPQGARGSEGPGQVR	191	

[illegible]


```
Db 456 GPKGAPGERGSPGLAGPKGANGDPGRPGEPGLFGARGLTGRPDGADGQKGVGSPGAGED 515
QY 298 GRPPGPPGARGQAGVMGFGPKGAAGRPGKAGRGVPGPCAVGPAKDGEGAQGGPP 357
Db 516 GRGPPGPGARGQAGVMGFGPKGANGRPGKAGRGVPGPCAVGPAKDGEGAQGGPP 575
QY 358 GPAGPAGERGEGAGSPGFGQGLGPGAPGPGGAGKPGGQGVPGDLGAPGSPGAGE--- 413
Db 576 GPAGPAGERGEGAGSPGFGQGLGPGAPGPGGAGKPGGQGVPGGAGAPGLVPGRGERGFP 635
QY 414 -----PGPTGLP-----GPPGERGPGSRGPPGADGVAGPKGPA 447
Db 636 GERGSPGAGQGLQPRGLTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 695
QY 448 GERGS-----PGPAGPKGSPGAGRPGGEGAGLPGAKGLTGP 483
Db 696 GDRGDVGEKPGAPGKDGGRGLTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 755
QY 484 GSPGPDGKTGPPGAGQGRGPPGPPGARGQAGVMGFGPKGAAGRPGKAGRGVPGPCAVGPA 534
Db 756 GETGPGTSGIAGPPGADGQPKAGKEGQGEAGQKDGAGAPGQPSGAPGPGGPTGVTGPK 815
QY 535 GERGVGPPGA-----VGPAGKDGEGAQAGPPGPA---GPAGERGQAGSPGFQ 582
Db 816 GARGAQQPPGATGPPGAGRGVPGPSNGNPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 875
QY 583 GLPGPAGPPGAGKPGGQGVPGDLGAPGSPGAGEPGLTGLPGERGPGSRGPPGAD 642
Db 876 GLQGPAGPPGKPGFGDGSAGRPPGQGLAGQGIIVGLPQGRGERGPPGLPGPSGEP 935
QY 643 GVAGPKGPPAGSRGSPGAPGKSGPGEAGRPGGAGLPCA-----KGLT 684
Db 936 GQCGAPGASGDRGPPGVPGLTGPAGEPGRGSPGADGPPGRDGAAGVKGRDGTGAV 995
QY 685 GSPSPGPDGKTGPPGP-----AGQDRGPPGPPGARGQAGVMGFGPKGAAGRP 735
Db 996 GAPGAPGPPSPGAPGTGKQDRGEAGAQGPMGSPGAPGARGIQGPQGRGDKGAGEP 1055
QY 736 GK---ACERGVPGPAGVAGKDGEGAQAGPPGAPGAPGAGERGQAGSPGFGQLGPGA 792
Db 1056 GERGLKHRTFTGLQGLPFPSPGQAGSPGPRGPPGVPFSGKDGAMGTPGI 1115
QY 793 GPGEAGKPGGQGVPGDLGAPGSPGAG 820
Db 1116 GPPGPRGSGETGAPGPPGPPGPPGPPG 1143

RESULT 15
US-09-500-811-20
; Sequence 20, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Oviast, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
```

```
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
; US-09-500-811-20

Query Match 53.4%; Score 2476; DB 3; Length 1418;
Best Local Similarity 50.9%; Pred. No. 1.4e-152;
Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;

QY 1 GPPGEPGPTGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60
Db 156 GNPGEPEGVSGPMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 215
QY 61 -----GAGLPGAKGLTGP-----GSPGP-----DGKTGPPGPA--- 90
Db 216 GVKHGRGYPGLDGAKGAGAGPVKGSGSPGNGSPGPMGPPGLPGLRGRGTGPAAGAR 275
QY 91 GQDRPPGPPPPGARGQAGVMGFP-----GPKGAAGRPGKAGRGVPGPPGAVGPA 141
Db 276 GNDQPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 335
QY 142 -----CKGEGAQAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 195
Db 336 GASGNPCTDGIKAGKSAGAPGIAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 395
QY 196 GD---LGAPGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 237
Db 396 GEQGPKEGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 455
QY 238 GPKGAPAGERGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 297
Db 456 GPKGAPAGERGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 515
QY 298 GRPPGPPGPPGARGQAGVMGFGPKGANGRPGKAGRGVPGPCAVGPAKDGEGAQGGPP 357
Db 516 GRGPPGPPGARGQAGVMGFGPKGANGRPGKAGRGVPGPCAVGPAKDGEGAQGGPP 575
QY 358 GPAGPAGERGEGAGSPGFGQGLGPGAPGPGGAGKPGGQGVPGDLGAPGSPGAGE--- 413
Db 576 GPAGPAGERGEGAGSPGFGQGLGPGAPGPGGAGKPGGQGVPGGAGAPGLVPGRGERGFP 635
QY 414 -----PGPTGLP-----GPPGERGPGSRGPPGADGVAGPKGPA 447
Db 636 GERGSPGAGQGLQPRGLTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 695
QY 448 GERGS-----PGPAGPKGSPGAGRPGGEGAGLPGAKGLTGP 483
Db 696 GDRGDVGEKPGAPGKDGGRGLTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 755
QY 484 GSPGPDGKTGPPGAGQGRGPPGPPGARGQAGVMGFGPKGAAGRPGKAGRGVPGPCAVGPA 534
Db 756 GETGPGTSGIAGPPGADGQPKAGKEGQGEAGQKDGAGAPGQPSGAPGPPGPTGVTGPK 815
```


Qy	535	GERGVGPPGA-----VBPAGKDGAGAGQPPGPA---GPAGERGEQGPAGSPGFQ	582
Db	816	GARGAQPPGATGTPGAAGRVPFGSGNPNPFPFPGSGKDGPKGARGDSGPPGRAGEP	875
Qy	593	GLPGPAGPPGAGKPGQGVPCDLGAPGSPGAPGPGTGLPGPPGERGPGSRGFFQAD	642
Db	876	GLQGPAGPPGKGBPGDDGPGSAGPPGFQGLAGQGI VGLPGQGERGFFGLPGPSGEP	935
Qy	643	GVAGPKGPAGERSGSPGAPGPKGSPGEAGRPGEAGLPGA-----KGLT	684
Db	936	GOQAGAPGASGDRGPPGVPFGLTGPAGEPRGREGSPGADGPPGRDGAAGVKGDRGETGAV	995
Qy	685	GSPGSPGPDGKTGPPGP-----AGODRPGPPPGPGARGOAGVMGFPKGAAGEP	735
Db	996	CAPGAPGPPGSPGAPGTPGKQDRGEAGAGQPMGSPGAPAGITQPGQPGDKGEAGEP	1055
Qy	736	GK---AGERGVGPPGAVGPPAGKDGAGAGAGQPPGPGAPAGERGQGPAGSPGQGLPGPA	792
Db	1056	GERGLKGRHGTGLQGLPGPPGSDQAGSPGSPGPRGPPGVGPGSKDGANGIPGI	1115
Qy	793	GPPGEAGKPGEGQVPCDLGAPGSPGAG	820
Db	1116	GPPGRGRSGETGPAGPPGPNPGPPG	1143

Search completed: June 17, 2005, 15:17:00
 Job time : 41.8485 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 15:02:42 ; Search time 89.9315 Seconds
(without alignments)
3505.413 Million cell updates/sec

Title: US-10-658-989A-3
Perfect score: 4640
Sequence: 1 GPPGEPGTGLPGERGG.....GEQVPGDLGAPGSPGAGG 821

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1714042 seqs, 38379560 residues

Total number of hits satisfying chosen parameters: 1714042

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4640	100.0	821	17	US-10-658-989A-3
2	3482	75.0	617	17	US-10-658-989A-2
3	2769	59.7	1014	17	US-10-901-816A-6
4	2769	59.7	1014	17	US-10-901-816A-10
5	2767	59.6	1014	17	US-10-901-816A-9
6	2767	59.6	1014	17	US-10-901-816A-11
7	2765.5	59.6	1449	15	US-10-402-089-8
8	2765.5	59.6	1449	15	US-10-402-072A-8
9	2765	59.6	1014	17	US-10-901-816A-5
10	2765	59.6	1057	15	US-10-104-889-20
11	2765	59.6	1107	15	US-10-104-889-11
					Sequence 3, Appli
					Sequence 2, Appli
					Sequence 6, Appli
					Sequence 10, Appli
					Sequence 9, Appli
					Sequence 11, Appli
					Sequence 8, Appli
					Sequence 5, Appli
					Sequence 20, Appli
					Sequence 11, Appli

12	2765	59.6	1171	15	US-10-104-889-8	Sequence 8, Appli
13	2765	59.6	1388	15	US-10-104-889-10	Sequence 10, Appli
14	2765	59.6	1464	14	US-10-216-705-21	Sequence 21, Appli
15	2765	59.6	1464	16	US-10-357-851-1	Sequence 1, Appli
16	2765	59.6	1464	16	US-10-358-024-1	Sequence 1, Appli
17	2765	59.6	1464	16	US-10-788-792-150	Sequence 150, App
18	2761	59.5	1461	16	US-10-468-091-25	Sequence 25, Appli
19	2761	59.5	1464	15	US-10-291-265-243	Sequence 243, App
20	2759	59.5	1057	15	US-10-104-889-16	Sequence 16, Appli
21	2759	59.4	1014	17	US-10-901-816A-7	Sequence 7, Appli
22	2755	59.4	1464	10	US-09-918-715-261	Sequence 261, App
23	2755	59.4	1464	14	US-10-060-036-159	Sequence 159, App
24	2755	59.4	1464	14	US-10-171-311-36	Sequence 36, Appli
25	2755	59.4	1464	14	US-10-149-352-2	Sequence 2, Appli
26	2755	59.4	1464	14	US-10-177-293-65	Sequence 55, Appli
27	2755	59.4	1464	14	US-10-301-822-28	Sequence 79, Appli
28	2755	59.4	1464	16	US-10-734-564-79	Sequence 261, App
29	2755	59.4	1464	16	US-10-474-794-261	Sequence 2289, Ap
30	2755	59.4	1464	16	US-10-723-860-2289	Sequence 157, App
31	2755	59.4	1464	17	US-10-852-335A-157	Sequence 8, Appli
32	2754	59.4	1014	17	US-10-901-816A-8	Sequence 2, Appli
33	2752.5	59.3	1463	15	US-10-402-089-2	Sequence 13, Appli
34	2752.5	59.3	1463	15	US-10-402-072A-2	Sequence 6, Appli
35	2752	59.3	1014	17	US-10-901-816A-13	Sequence 26, Appli
36	2750	59.3	1169	15	US-10-104-889-6	Sequence 12, Appli
37	2732.5	58.9	1453	16	US-10-468-091-26	Sequence 18, Appli
38	2729	58.8	1014	17	US-10-901-816A-12	Sequence 49, Appli
39	2704	58.3	1341	14	US-10-058-124-18	Sequence 10, Appli
40	2511	54.1	822	15	US-10-342-331-49	Sequence 1, Appli
41	2488	53.6	1017	16	US-10-639-286-10	Sequence 20, Appli
42	2479	53.4	1014	14	US-10-194-441A-1	Sequence 5, Appli
43	2476	53.4	1418	14	US-10-058-124-20	Sequence 48, Appli
44	2476	53.4	1418	16	US-10-468-091-5	
45	2460	53.0	1014	14	US-10-194-441A-48	

ALIGNMENTS

RESULT 1
US-10-658-989A-3
; Sequence 3, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-4
US-10-658-989A-3

Query Match	100.0%	Score 4640;	DB 17;	Length 821;
Best Local Similarity	100.0%	Pred. No. 2.9e-225;		
Matches 821;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GPPGEPGTGLPGERGGPGAGVAGPKGSPGAGPKGSPGAGRP	60	
Db	1	GPPGEPGTGLPGERGGPGAGVAGPKGSPGAGPKGSPGAGRP	60	
Qy	61	GEAGLPGAKGLTGTSGSPGPGKGTGPPGAGQDGPDPGPPGARGQGVGFFGPKGAA	120	

Db 61 GEAGLPGAKGLTSGSPGPPGDKTTPPGPAGQDGRPGPPPGARGQAGVMGFPKGA 120
QY 121 GEPGKAGRGVPPGAVGPAKDGCAAGQCPGCPAGPAGRGQAGSPGFGQLPGPA 180
Db 121 GEPGKAGRGVPPGAVGPAKDGCAAGQCPGCPAGPAGRGQAGSPGFGQLPGPA 180
QY 181 GPPGAGKPGGQGVPGDGLGAPGPGAPGEPGPTGLPGPPGGRGGPGSRGFGADGAGPK 240
Db 181 GPPGAGKPGGQGVPGDGLGAPGPGAPGEPGPTGLPGPPGGRGGPGSRGFGADGAGPK 240
QY 241 GPAGRGSPGAPGKSGPGEAGRCAGLPGCAKGLTSGSPGPPGDKTTPPGPAGQDGRP 300
Db 241 GPAGRGSPGAPGKSGPGEAGRCAGLPGCAKGLTSGSPGPPGDKTTPPGPAGQDGRP 300
QY 301 GPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVPPGAVGPAKDGAGAGQPPGPA 360
Db 301 GPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVPPGAVGPAKDGAGAGQPPGPA 360
QY 361 GPAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 420
Db 361 GPAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 420
QY 421 GPPGGRGPGSRGFPFGADGAGVAGPKGAGRGSPGAPGKSGPGEAGRPGEAGLPGAKGLT 480
Db 421 GPPGGRGPGSRGFPFGADGAGVAGPKGAGRGSPGAPGKSGPGEAGRPGEAGLPGAKGLT 480
QY 481 GSPGSPGPDGKTGTPPGAGQDGRPGPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVVP 540
Db 481 GSPGSPGPDGKTGTPPGAGQDGRPGPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVVP 540
QY 541 GPPGAVGPAKDGAGAGQPPGAPGAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 600
Db 541 GPPGAVGPAKDGAGAGQPPGAPGAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 600
QY 601 GVPDGLGAPGSPGAPGFPPTGLPGPPGGRGGPGSRGFGADGAGVAGPKGAGRGSPGPA 660
Db 601 GVPDGLGAPGSPGAPGFPPTGLPGPPGGRGGPGSRGFGADGAGVAGPKGAGRGSPGPA 660
QY 661 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPPGDKTTPPGPAGQDGRPGPPGARGQAG 720
Db 661 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPPGDKTTPPGPAGQDGRPGPPGARGQAG 720
QY 721 GVMGFPKGAAGEPKKAGRGVPPGAVGPAKDGCAAGQCPGCPAGPAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 780
Db 721 GVMGFPKGAAGEPKKAGRGVPPGAVGPAKDGCAAGQCPGCPAGPAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 780
QY 781 GSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPG 821
Db 781 GSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPG 821

RESULT 2
US-10-658-989A-2
; Sequence 2, Application US/1065989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-3

US-10-658-989A-2

Query Match 75.0%; Score 3482; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 2.5e-167;
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPPGEGPTGLPGPPGGRGGPGSRGFGADGAGVAGPKGAGRGSPGAPGKSGPGEAGRP 60
Db 1 GPPGEGPTGLPGPPGGRGGPGSRGFGADGAGVAGPKGAGRGSPGAPGKSGPGEAGRP 60
QY 61 GEAGLPGAKGLTSGSPGPPGDKTTPPGPAGQDGRPGPPPGARGQAGVMGFPKGA 120
Db 61 GEAGLPGAKGLTSGSPGPPGDKTTPPGPAGQDGRPGPPPGARGQAGVMGFPKGA 120
QY 121 GEPGKAGRGVPPGAVGPAKDGCAAGQCPGCPAGPAGRGQAGSPGFGQLPGPA 180
Db 121 GEPGKAGRGVPPGAVGPAKDGCAAGQCPGCPAGPAGRGQAGSPGFGQLPGPA 180
QY 181 GPPGAGKPGGQGVPGDGLGAPGPGAPGEPGPTGLPGPPGGRGGPGSRGFGADGAGPK 240
Db 181 GPPGAGKPGGQGVPGDGLGAPGPGAPGEPGPTGLPGPPGGRGGPGSRGFGADGAGPK 240
QY 241 GPAGRGSPGAPGKSGPGEAGRCAGLPGCAKGLTSGSPGPPGDKTTPPGPAGQDGRP 300
Db 241 GPAGRGSPGAPGKSGPGEAGRCAGLPGCAKGLTSGSPGPPGDKTTPPGPAGQDGRP 300
QY 301 GPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVPPGAVGPAKDGAGAGQPPGPA 360
Db 301 GPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVPPGAVGPAKDGAGAGQPPGPA 360
QY 361 GPAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 420
Db 361 GPAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 420
QY 421 GPPGGRGPGSRGFPFGADGAGVAGPKGAGRGSPGAPGKSGPGEAGRPGEAGLPGAKGLT 480
Db 421 GPPGGRGPGSRGFPFGADGAGVAGPKGAGRGSPGAPGKSGPGEAGRPGEAGLPGAKGLT 480
QY 481 GSPGSPGPDGKTGTPPGAGQDGRPGPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVVP 540
Db 481 GSPGSPGPDGKTGTPPGAGQDGRPGPPGPPGARGQAGVMGFPKGAAGEPKKAGRGVVP 540
QY 541 GPPGAVGPAKDGAGAGQPPGAPGAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 600
Db 541 GPPGAVGPAKDGAGAGQPPGAPGAGRGQAGSPGFGQLPGPAGPGEAGKPGQGVPGDGLGAPGSPGAPGFPPTGLP 600
QY 601 GVPDGLGAPGSPGAPGFPPTGLPGPPGGRGGPGSRGFGADGAGVAGPKGAGRGSPGPA 616
Db 601 GVPDGLGAPGSPGAPGFPPTGLPGPPGGRGGPGSRGFGADGAGVAGPKGAGRGSPGPA 616

RESULT 3
US-10-901-816A-6
; Sequence 6, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-6

Query Match 59.7%; Score 2769; DB 17; Length 1014;
Best Local Similarity 55.5%; Pred. No. 1.9e-131;
Matches 556; Conservative 33; Mismatches 230; Indels 182; Gaps 17;

QY 1 GPPGEPGTGLPGPPGRRGGPGSGRFP-----GADGVAGPKGPAGRGSGPG 47
DB 13 GPPGAPGQGGPPGEGFEGASG-PWMPGPPGPPGKNGDDGEAGKPGRGSGPG 71
QY 48 AGPKGSPGAGRPGEAGLPGAKGLTSGSPGPGDKTTPGPAGQDGRPGPPGARGQ 107
DB 72 QCARGLPCTAGLPGMKHGRGFSGLDGAAGDAGPAGKGPSPGEGAPGQMGPPGLPE 131
QY 108 AGVWGFPGKGAAGEPGKAGRGVPPGAVGP-----AGKDEAGAGQPPGAPAGE 161
DB 132 RGRPGAPGARGNDGATGAAGPPGPTGAPGPPGFAVGAKGAGPQGRSGEPGV 191
QY 162 RGEQGP-----AGSPGFQGLPGPAGPPGAGKPG 191
DB 192 RGEFPGPPAGNPGADGQCGKANGAPGIAAGPFFGARGSGPGGPPGP 251
QY 192 QGVPGDLGAP-----GPSGPA-----GEPGPTGLPGPPGRRG 224
DB 252 KGNSEPGAGSGKDTGAKGPPGVGPPGCPAGEEGKRGARGEPGPTGLPGPPGRRG 311
QY 225 PGRGFPAGDGVAGPKGAGRGSGPAGPKGSPGEGACRPGEAGLPGAKGLTSGSPGP 284
DB 312 PGRGFPAGDGVAGPKGAGRGSGPAGPKGSPGEGACRPGEAGLPGAKGLTSGSPGP 371
QY 285 DGKTGPPGAGQDGRPGPPGARGQAGVWGFPGKGAAGEPGKAGRGVPPGAVGP 344
DB 372 DGKTGPPGAGQDGRPGPPGARGQAGVWGFPGKGAAGEPGKAGRGVPPGAVGP 431
QY 345 AGKDEAGAGQPPGAPGAGRGSGPAGSGPFGGLPGPAGPPGAGKPGSQVPGDLGA 404
DB 432 AGKDEAGAGQPPGAPGAGRGSGPAGSGPFGGLPGPAGPPGAGKPGSQVPGDLGA 491
QY 405 PGPSGPAGE-----PGP-----TGLPGPPGRRGSGRGP 434
DB 492 PGPSGARGRFPGRRGVPQPPGPPRGANGAFNGDKAGDAGAPAGSQAGPLQGM 551
QY 435 PGADGVAGPKGAGRGSGPAGPKGSPG-----AGRPGGAGLPGAK-----477
DB 552 PGERGAAGLPGKDRDAGPKADGSGKDGVRGLTGPDPGPPAGAPGDKGSGSGP 611
QY 478 -GLTSGSPGPGDKTTPGPAGQDGRPGPPGARGQ-----AGVMGF 521
DB 612 AGPTGARGAGCDRGEPGPPGAPGAGPPGADGQPGKAGGECDAKGDAGPPGAPG 671
QY 522 PGPKGAAGEPGKAGRGVPPGPPGA-----VGPAGKDEAGAGQPPGAPG---AGE 569

RESULT 4
US-10-901-816A-10
; Sequence 10, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; PRIOR FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-10

Query Match 59.7%; Score 2769; DB 17; Length 1014;
Best Local Similarity 55.5%; Pred. No. 1.9e-131;
Matches 556; Conservative 33; Mismatches 230; Indels 182; Gaps 17;

QY 1 GPPGEPGTGLPGPPGRRGGPGSGRFP-----GADGVAGPKGPAGRGSGPG 47
DB 13 GPPGAPGQGGPPGEGFEGASG-PWMPGPPGPPGKNGDDGEAGKPGRGSGPG 71
QY 48 AGPKGSPGAGRPGEAGLPGAKGLTSGSPGPGDKTTPGPAGQDGRPGPPGARGQ 107
DB 72 QCARGLPCTAGLPGMKHGRGFSGLDGAAGDAGPAGKGPSPGEGAPGQMGPPGLPE 131
QY 108 AGVWGFPGKGAAGEPGKAGRGVPPGAVGP-----AGKDEAGAGQPPGAPAGE 161
DB 132 RGRPGAPGARGNDGATGAAGPPGPTGAPGPPGFAVGAKGAGPQGRSGEPGV 191
QY 162 RGEQGP-----AGSPGFQGLPGPAGPPGAGKPG 191
DB 192 RGEFPGPPAGNPGADGQCGKANGAPGIAAGPFFGARGSGPGGPPGP 251
QY 192 QGVPGDLGAP-----GPSGPA-----GEPGPTGLPGPPGRRG 224
DB 252 KGNSEPGAGSGKDTGAKGPPGVGPPGCPAGEEGKRGARGEPGPTGLPGPPGRRG 311
QY 225 PGRGFPAGDGVAGPKGAGRGSGPAGPKGSPGEGACRPGEAGLPGAKGLTSGSPGP 284
DB 312 PGRGFPAGDGVAGPKGAGRGSGPAGPKGSPGEGACRPGEAGLPGAKGLTSGSPGP 371
QY 285 DGKTGPPGAGQDGRPGPPGARGQAGVWGFPGKGAAGEPGKAGRGVPPGAVGP 344
DB 372 DGKTGPPGAGQDGRPGPPGARGQAGVWGFPGKGAAGEPGKAGRGVPPGAVGP 431
QY 345 AGKDEAGAGQPPGAPGAGRGSGPAGSGPFGGLPGPAGPPGAGKPGSQVPGDLGA 404
DB 432 AGKDEAGAGQPPGAPGAGRGSGPAGSGPFGGLPGPAGPPGAGKPGSQVPGDLGA 491
QY 405 PGPSGPAGE-----PGP-----TGLPGPPGRRGSGRGP 434
DB 492 PGPSGARGRFPGRRGVPQPPGPPRGANGAFNGDKAGDAGAPAGSQAGPLQGM 551
QY 435 PGADGVAGPKGAGRGSGPAGPKGSPG-----AGRPGGAGLPGAK-----477
DB 552 PGERGAAGLPGKDRDAGPKADGSGKDGVRGLTGPDPGPPAGAPGDKGSGSGP 611
QY 478 -GLTSGSPGPGDKTTPGPAGQDGRPGPPGARGQ-----AGVMGF 521
DB 612 AGPTGARGAGCDRGEPGPPGAPGAGPPGADGQPGKAGGECDAKGDAGPPGAPG 671
QY 522 PGPKGAAGEPGKAGRGVPPGPPGA-----VGPAGKDEAGAGQPPGAPG---AGE 569


```

Db      672  PGTGPPGAGKAGAGAGGATGTFGGAAGRVGPPGPGSNAGPPGPGFAGKEGKGP 731
Qy      570  RGEQGPAGSPGQGLGPPAGPAGGAGKPGEGQGVGDLGAPGSPGAGEPPTGLKGPGE 629
Db      732  RGETGPAGRPVEGPPGPPGAGKEGSGADGAPAGTGTGPGQIAGQGVVGLPGQGE 791
Qy      630  RGGPSRRPGADGVAGPKGAPAGBRGSPGAPG---XGSPGEAGRPGEAGLPGAKGLTGS 686
Db      792  RGFTGLPFGSPGPKQKQSGASGERGPPGMPGPGLAGPPGEGSREGAPGAGESGFRDGS 851
Qy      687  PGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVGMGFPKGAAGEPKKAGRGVPGP 746
Db      852  PKAGKDRGETGPPGAPGAPGAPGVPVGPAGKSGDRGETGTPAGPAGVPVGPAGARGPAG 911
Qy      747  PGAVGPAGKQGE-----AGAQQPPGP-----AGPAGGERGEQGPAGS 782
Db      912  QGPRGDKGTGEGQDGRGKQHRGFSGLQGPFGPPGSPGEGQGPAGPRGPPGASGA 971
Qy      783  P---GFOGLPGAPPPGAGKPGEGQGVGDLGAPGSPGAG 820
Db      972  PGKDLNGLPGPIGPPGPRGRTGDAGVPVGPVPPGPPGPPG 1012

```

RESULT 5

```

US-10-901-816A-9
; Sequence 9, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; PRIOR FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1014
; TYPE: PRP
; ORGANISM: Homo sapiens
US-10-901-816A-9

```

```

Query Match      59.6%; Score 2767; DB 17; Length 1014;
Best Local Similarity 55.3%; Pred. No. 2.4e-131;
Matches 554; Conservative 34; Mismatches 231; Indels 182; Gaps 16;

Qy      1  GPPGERGPTGLPDPGERGSGRGP-----GADGVAGPKPAGERSGPP 47
Db      13  GPPGAPGQPGQFQGPPEGEPKASG-PMGPPPPGPPGKNGDDGEAGKPRGERGPPGP 71
Qy      48  AGPKGSPGAGRPGEAGLPGAKGLTGSPPGPDGKTGPPGAGQDGRPPGPPGARGQ 107
Db      72  QGARGLPTAGLPMKQHRGFSGLDGAADGAPGPKGPPGSPGNGAPQMGPPFLPGE 131
Qy      108  AGVMGPPGKGAAGEPKAGRGVGFPPGAVGP-----AGKDGAGAGQGGPPGAPAGE 161
Db      132  RGRPGAPGAGARGNDGATGAAGPPTGTPAGPPGPPGAVAGKAGAGQGPFGSGPQGV 191
Qy      162  RGEQGP-----AGSPGQGLPAGPAGPGEAGKPG 191
Db      192  RGEPPGPPGAGAGPAGNPGADQGPAGKANGAPGAGPAGPAGPAGPAGPAGP 251
Qy      192  QGVPGDLGAP-----GPSGA-----GPPGPTGLPAGPGERGG 224
Db      252  KNSGEPGAPGSKDGTAKGPGVGVQFPAGPAGGKRGARGEPGTGLPAGPGERGG 311
Qy      225  PGRGPPGADGVAGPKPAGERSGPPGAPKSPGEGRPGAGLPGAKGLTGSFGSPG 284

```

```

Db      312  PGRGPPGADGVAGPKPAGERSGPPGAPKSGPGEAGRPGEAGLPGAKGLTGSFGSPG 371
Qy      285  DGKTGPPGAGQDGRPPGPPGARGQAGVGMGFPKGAAGBPGKAGRGVFPAGAVGP 344
Db      372  DGKTGPPGAGQDGRPPGPPGARGQAGVGMGFPKGAAGBPGKAGRGVFPAGAVGP 431
Qy      345  AGKDGAGAGQGGPPGAPAGERGEQGPAGSPGQGLGPPGAPGPPGAGKPGEGVPGDLGA 404
Db      432  AGKDGAGAGQGGPPGAPAGERGEQGPAGSPGQGLGPPGAPGPPGAGKPGEGVPGDLGA 491
Qy      405  PGPSPGAGE-----PGP-----TGLPGPPGPPGPPGSGRGP 434
Db      492  PGPSPGARGERPGBERGVQGPAGPRGANGAPGNDGAKGADGAPGAPGSGAGFLQGM 551
Qy      435  PGADGVAGPKPAGERSGPPGAPKSGPGE-----AGRPGEAGLPGA 476
Db      552  PGERGAALPGPKGDRGADGPKGADGSPGKDVRLGTGPIGPPGAPGAPGPPGSGSPG 611
Qy      477  KGLTSPGSPGPDGKTGTPGPPGAGQDGRPPGPPGARGQ-----AGVMGF 521
Db      612  AGPTGARGAPGDRGEPGPPGPPGAGPAGPAGQGPAGKPGDAGKAGDAGPPGAPGAP 671
Qy      522  PGPKGAAGEPKKAGRGVPPGPPGA-----VGPAGKDGAGAGQGGPPGAPG---AGE 569
Db      672  PGP1GPPGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPPGPPGPPGPPGPPGPPG 731
Qy      570  RGEQGPAGSPGQGLGPPAGPAGGAGKPGEGQGVGDLGAPGSPGAPGPPGTGLPAGP 629
Db      732  RGETGPAGRPVEGPPGPPGAGKSGADGAPAGTGTGPGQIAGQGVVGLPGQGE 791
Qy      630  RGGPSRRPGADGVAGPKGAPAGBRGSPGAPG---XGSPGEAGRPGEAGLPGAKGLTGS 686
Db      792  RGFTGLPFGSPGPKQKQSGASGERGPPGMPGPPGLAGPPGEGSREGAPGAGESGFRDGS 851
Qy      687  PGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVGMGFPKGAAGEPKKAGRGVVP 746
Db      852  PKAGKDRGETGPPGAPGAPGAPGVPVGPAGKSGDRGETGTPAGPAGVPVGPAGARGPAG 911
Qy      747  PGAVGPAGKQGE-----AGAQQPPGP-----AGPAGGERGEQGPAGS 782
Db      912  QGPRGDKGTGEGQDGRGKQHRGFSGLQGPFGPPGSPGEGQGPAGPRGPPGASGA 971
Qy      783  P---GFOGLPGAPPPGAGKPGEGQGVGDLGAPGSPGAG 820
Db      972  PGKDLNGLPGPIGPPGPRGRTGDAGVPVGPVPPGPPGPPG 1012

```

RESULT 6

```

US-10-901-816A-11
; Sequence 11, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; PRIOR FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1014
; TYPE: PRP
; ORGANISM: Homo sapiens
US-10-901-816A-11

```

```

Query Match      59.6%; Score 2767; DB 17; Length 1014;
Best Local Similarity 55.3%; Pred. No. 2.4e-131;
Matches 554; Conservative 34; Mismatches 231; Indels 182; Gaps 16;

```



```

QY 1 GPPGPGTGLPGRGGRGSGRGP-----GADGAGKPGAGRGSGRP 47
DB 13 GPPGAPGQGGPGEPEGASG-PMGPPGPPGPKNGDGDGAGKPGRGERGPPG 71
QY 48 AGPKSGEAGRPAGLPGAKGLTSGSPGDPCKTGPAGDGRPPGPPGARGQ 107
DB 72 QGARGLPTAGLPGMKHGRGSLDAGKADGAPGKPSGPGENGAPGQMGPPGLGE 131
QY 108 AGVMGPPGKGAAGPFGKAGRGVPPGAVGP-----AGKDGAGAGQGGPPGAPAGE 161
DB 132 RGRPCAPGARGNDGATGAAGPPGPTGAPGPPGAVGAKGAGPQGRSGPQGV 191
QY 162 RGEQGP-----AGSPGQGLPGAPPPGEGAGKGE 191
DB 192 RGEPPGPGAGAGPAGNPGADGQPGKAGANGAPGPPGARGSPGQGGPPGP 251
QY 192 QGVPGDLGAP-----GPSGPA-----GPPGPTGLPGRGGRGG 224
DB 252 KNSGEPGAPGSKGDTGAKGPPGVGPPGAPGEEKRGARGEPGPTGLPGRGGRGG 311
QY 225 PGSRGFPAGDGVAGPKPAGRGSGPPAGPKSPGEGAGRPGEAGLPGAKGLTSGSPG 284
DB 312 PGSRGFPAGDGVAGPKPAGRGSGPPAGPKSPGEGAGRPGEAGLPGAKGLTSGSPG 371
QY 285 DKTGPPGAGDGRPPGPPGARGQAGVMGPPGPKGAAGHPGKAGRGVPPGAVGP 344
DB 372 DKTGPPGAGDGRPPGPPGARGQAGVMGPPGPKGAAGHPGKAGRGVPPGAVGP 431
QY 345 AGKDGAGAGQPPGAGPAGRGEGOGPAGSPGQGLPGAPGPPGAGKPGGQVPGDLGA 404
DB 432 AGKDGAGAGQPPGAGPAGRGEGOGPAGSPGQGLPGAPGPPGAGKPGGQVPGDLGA 491
QY 405 PGPSGPA-----PGP-----TGLPGRGGRGSGRGP 434
DB 492 PGPSGARGRPPGGRGVQPPGPPAGRGANGAPGNDGAKGDAGAPGAGSGQAPGLQGM 551
QY 435 PGADGVAGPKPAGRGSGPPAGPKSGE-----AGRPGEAGLPGA 476
DB 552 PGERGAAGLPGPKGRDAGPKGADGSPGKQGVRLTGPIGPPGAGAPGPPGSGSGSP 611
QY 477 KGLTSGSPGDPKTPGPRGAGQGRGPPGPPGARGQ-----AGVMGF 521
DB 612 AGPTGARGAPGRGPPGPPGAGFAGPPGADQPGKAGGPDGAKGDAGPPGAPGAP 671
QY 522 PGPKAAGBPAGRGVPPGPA-----VGPAGKDGAGAGQGGPPGAPG---AGE 569
DB 672 PGIENVGAPGAKGARGSGPPGATGFPAGAGRVGPPGPPSGNAGPPGPPGAGKGGKGP 731
QY 570 RGEQPGSPGQGLPGAPGPPGAGKPGGQGVPGDLGAPGSPGAPGPTGLPGRGPE 629
DB 732 RGETGAPRGVBVPPGPPGAGKSGPCADGAPAGAPGTPGQGIAGQGVVGLPGQGE 791
QY 630 RGGPSRGPFGADGVAGPKPAGRGSGPPAGP-----KSGPGEAGRPGEAGLPGAKGLTGS 686
DB 792 RGFPLGPPSGEPGKQGSAGSGRGP PGMPGPPGLAGPPGSSGREGAPGAGSGSPRDS 851
QY 687 PGSPGDKTGPAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGHPGKAGRGVGP 746
DB 852 PGAKGDRGTGAPGPPGAPGAPGPPGAPGAGKSGDRGTGAPGAPGPPGAPGAPGAP 911
QY 747 PGAVGAPGKDG-----AGAGQPPGP-----AGPAGERGEGPAGS 782
DB 912 QGPRGDKGTGQDRGIKGRGFSGLQGGPPGPPGSGQPSGASGAPGPPGSGAGA 971
QY 783 P---GFGGLPGAPGPPGAGKPGGQGVPGDLGAPGSGPAG 820
DB 972 PGKDLGLGLPGIIGPPGRGRTGDAGVPPGPPGPPGPP 1012

```

RESULT 7

US-10-402-089-8

; Sequence 8, Application US/10402089

```

; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: F0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 1449
; TYPE: PRT
; ORGANISM: Sus scrofa
; US-10-402-089-8

Query Match 59.6%; Score 2765.5; DB 15; Length 1449;
Best Local Similarity 56.7%; Pred. No. 3.7e-131;
Matches 559; Conservative 32; Mismatches 226; Indels 169; Gaps 21;

QY 1 GPPGPGTGLPGRGGRGSGRGP-----RGFPGA-----DGVAGPKPAGRGSGP 48
DB 193 GPPGAPGQGGPGEPEGASGPMGPPGPPGPKNGDGDGAGKPGRGERGPPG 252
QY 49 GPKSGPAGRGP-----GEAGLPGAKGLTSGSPGDPCKTGPAGQD 93
DB 253 GARGLPTAGLPGMKHGRGSLDAGKADGAPGPKPSGPGENGAPGQMGPPGLG 312
QY 94 GRPGRPPGARGQ-----AGVMGPPGPKGAAG-----PGKAGRGVPPGAVGAPGKDG 147
DB 313 GRPGRPPGARGNDGATGAAGPPGPTGAPGPPGAVGAKGAGPQGRSGPQGV 372
QY 148 GAQGGPPGAPGARGEGGP-----AGSPGQGLPGAPGPPGAGKPGG 192
DB 373 GEPGPPGAGAGPAGNPGADGQPGKANGAPGATAGAPGPPGARGSPGQGGPPG 432
QY 193 GVPDGLGAP-----GPSGPA-----GEPGPTGLPGRGGRGP 225
DB 433 GNSGEPGAPGSKGDTGAKGPPGTGVQPPGPPGAGEEKGARGEPGAPGLPGR 492
QY 226 GSRGFPAGDGVAGPKPAGRGSGPPAGPKSGPGEAGLPGAKGLTSGSPGPD 285
DB 493 GSRGFPAGDGVAGPKPAGRGSGPPAGPKSGPGEAGLPGAKGLTSGSPGPD 552
QY 286 KTGPPGAGDGRPPGPPGARGQAGVMGPPGPKGAAGHPGKAGRGVPPGAVG 345
DB 553 KTGPPGAGDGRPPGPPGARGQAGVMGPPGPKGAAGHPGKAGRGVPPGAVG 612
QY 346 KDGAGAGQPPGAPGARGEGOGPAGSPGQGLPGAPGPPGAGKPGGQVPGDLG 405
DB 613 KDGAGAGQPPGAPGARGEGOGPAGSPGQGLPGAPGPPGAGKPGGQVPGDLG 672
QY 406 GPSGPA-----PGP-----TGLPGRGGRGSGRGP 435
DB 673 GPSGARGRPPGGRGVQPPGPPAGRGANGAPGNDGAKGDAGAPGAPGAGLQGM 732
QY 436 GADGVAGPKPAGRGSGPPAGPKSGPGE-----AGRPGEAGLPGAKGLTSG 486
DB 733 GERGAAGLPGPKGRDAGAPKADGAPGKDGVRGLTGPIGPPGAPGAPGDKGT 792
QY 487 GPDGKTGPAGQDGRGPP-----PGPCARGQAGVMGF-----PGPKGAAG 539
DB 793 GPTGARGAPGRGPPGPPGAPGPPGADQPGKAGGTPGPPGPIGSGVAPGPPG 852
QY 540 PGPPGA-----VGPAGKDGAGAGQPPGPPGAPAGE-----RGEQGPAG 587
DB 853 AGPPGATGPPGAGRGVPPGPPGSGNAGPPGPPGAPGKSGKGRGTGAPG 912
QY 588 AGPPGAGKPGGQGVPGDLGAPGSPGAPGEPGPTGLPGRGGRGSGRGP 647

```



```
QY 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTPPPAGQDGRPPGPPGAR 105
Db 61 GRPGERGPPGQAGRLPGTAGLPGMKHGRFSGLDGAKGAGPAGPKGPPGSGPENGAP 120
QY 106 QGAGVMGPPGPKGAAGEPGKAGER-----GVPPGPGAVGPA-----GKDGEAGAQ 150
Db 121 GQMGPRGLPGERGRPGAGPAGARGNDGATGAAGPPGTPGAGPPGPGAVGAKGEAGPQ 180
QY 151 GPPGAPGAGRGEGP-----AGSPGPGGLPGPA 180
Db 181 GPRGSEGGVGRGPPGPPGAGAGPAGNPGADGPGAKGANGAPGAGPAGPAGPAGP 240
QY 181 GPPGEAGKPGQGVPGDLGAP-----GPSGPA-----GEPGPT 213
Db 241 GPQGGGPPGPKGNSGPPGAGPSKGDGTAKGEPGPGVQGGPPGAGEGKRGAGEPGPT 300
QY 214 GLPDPGGRGPPGSGRFFGADGAGKPGPAGERSGPGAGPKGSPGAGRPGEAGLPGAK 273
Db 301 GLPDPGGRGPPGSGRFFGADGAGKPGPAGERSGPGAGPKGSPGAGRPGEAGLPGAK 360
QY 274 GLTSGPSGPPDGKTPGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 333
Db 361 GLTSGPSGPPDGKTPGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 420
QY 334 GVPDPGAVGAGKDGEGAGAGQPPGPPGAGRGEGQPGAGSGPFGQGLPGPAGPPGEAGKP 393
Db 421 GVPDPGAVGAGKDGEGAGAGQPPGPPGAGRGEGQPGAGSGPFGQGLPGPAGPPGEAGKP 480
QY 394 GEOGVPGDLGAPGSPGAGE-----PGP-----TGLRPP 423
Db 481 GEOGVPGDLGAPGSPGAGEGPFGERGVQPPGPPGARGANGAPGNDGAKGDAGAP 540
QY 424 GERGGPSRPPGADGAVGAPKPGAGERSGPPAGPKGSPGE-----AGRPGEAGLP 474
Db 541 GSQAGPLQGMGERGAGAGLPKPGDRDAGPKGADSGKDGVRGLTGPPIGPPGAPAG 600
QY 475 GAK-----GLTSGPSGPPDGKTPGPPGAGQDGRPPGPPGARGQ-----515
Db 601 GDKGESGPPGAPGTGARGAPGDRGEPGPPGPPGAGPAGPAGGADGPPGAKGEPGCDAGAKGDA 660
QY 516 -----AGVMGPPGPKGAAGEPGKAGERGVPPGPA-----VGPAGKDGEGAGQPP 561
Db 661 GPPGAPGAGPPGLGNVGAACAKGARGSPGATGPPGAGRVGPPGPPGSGNAGPPPP 720
QY 562 GPAGP---AGERGEGPAGSPGFGQLPGPAGPAGGAGKPGGQGVPGDILGAPGSPGAGEP 618
Db 721 GPAGKEGKGRGETGTPAGRGVEVPPGPPGAGEKGSFGADGAGAPGTTGPGQIAGOR 780
QY 619 GPTGLPDPGPPGPPGSGRFFGADGAVGAPKPGPAGERSGPPAGP---KGSPPGAGRPGEA 675
Db 781 GVVGLPQGRGERGPPGLPSPGEPGKQFSGASGERGPPGPPGLAGPPGSGREGAP 840
QY 676 GLPAGKGLTSGSPGPPGKTPGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEP 735
Db 841 GAEGSPGRDGSFGAKGDRGETGTPAGPPGAPGAPGVPVAGKSGDRGETGTPAGPAGPV 900
QY 736 GKAGERGVPPGPPGAVGPAKDGGE-----AGAQQPPGP-----AGPA 771
Db 901 GPAGARGPAGPQGRDGETGEGQDRIKIHGRFSGLGQPPGPPGSGEGQPSGASGPA 960
QY 772 GEREGQGPAGSP---GFGQLPGPAGPAGGAGKPGGQGVPGDILGAPGSPGAP 820
Db 961 GPRGPPGSAGAPGDKGLNGLPGPIGPPGPRGRTGDAGVPPGPPGPPGPPG 1012
```

RESULT 10

```
US-10-104-889-20
; Sequence 20, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
```

```
; BROKAW, JANE
; ZHANG, GUANGHUI
; FAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVERTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-104-889-20
```

```
Query Match 59.6%; Score 2765; DB 15; Length 1057;
Best Local Similarity 55.4%; Pred. No. 3.1e-131;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGEPGPTGLPDPGPPGPPGSGRFFGADGAGKPGPAGRGSPGAPGPKGSGPGEA---57
Db 18 GPMGSPGPRGLPDPGPPGAPGPGQFGQPPGEPGSGPMGPPGPPGPKNGDDGEGAKP 77
QY 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTPPPAGQDGRPPGPPGAR 105
Db 78 GRPGERGPPGQAGRLPGTAGLPGMKHGRFSGLDGAKGADGAPGPKGPPGSGENGAP 137
QY 106 QGAGVMGPPGPKGAAGEPGKAGER-----GVPPGPGAVGPA-----GKDGEAGAQ 150
Db 138 GQMGPRGLPGERGRPGAGPAGARGNDGATGAAGPPGTPGAGPPGPGAVGAKGEAGPQ 197
QY 151 GPPGAPGAGRGEGP-----AGSPGPGGLPGPA 180
Db 198 GPRGSEGGVGRGPPGPPGAGAGPAGNPGADGPPGAKGANGAPGAGPAGPAGPAGP 257
QY 181 GPPGEAGKPGQGVPGDLGAP-----GPSGPA-----GEPGPT 213
Db 258 GPQGGGPPGPKGNSGPPGAGPSKGDGTAKGEPGPGVQGGPPGAGEGKRGAGEPGPT 317
QY 214 GLPDPGGRGPPGSGRFFGADGAVGAPKPGPAGERSGPPAGPKGSPGAGRPGEAGLPGAK 273
Db 318 GLPDPGGRGPPGSGRFFGADGAVGAPKPGPAGERSGPPAGPKGSPGAGRPGEAGLPGAK 377
QY 274 GLTSGPSGPPDGKTPGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 333
Db 378 GLTSGPSGPPDGKTPGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 437
QY 334 GVPDPGAVGAPKDGEGAGAGQPPGPPGAGPAGRGEGQPGAGSGPFGQGLPGPAGPPGEAGKP 393
```


[illegible]

RESULT 12

US-10-104-889-8
; Sequence 8, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.

COMPUTER READABLE FORM:

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO. 8:

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS.

SEQUENCE CHARACTERISTICS:
 LENGTH: 1171 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:

1-889-8
SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Query Match	59.6%;	Score 2765;	DB 15;	Length 1171;
Best Local Similarity	55.4%;	Pred. No. 3.3e-131;		
Matches 561: Conservative	31: Mismatches	228: Indels	192: Gaps	19

[illegible]

151	QY	GPDPGAGPAGRGGQGP-----ACSPGEGQLPGPA	180
198	Db	GRGRGEGPGVGRGPGPPGPAAGPAGNFCADQPGAKGANGAPGAGPFGPAGRGPS	257
181	QY	GPDPGAGKPGGQGVPGDLGAP-----GPSGPA-----GSPGGT	213
258	Db	GPQGGPGPPGKNSGPGAPGSKGDTGAKGEGPCPVGVQCPGAGEGKRGARGEGPT	317
214	QY	GLUPPPGRGPGFGRGPPPGADVAGPKPPAGRGSGPCGAPGKSGPGBAORPGBAGLPGAK	273
318	Db	GLUPPPGRGPGFGRGPPPGADVAGPKPPAGRGSGPCGAPGKSGPGBAORPGBAGLPGAK	377
274	QY	GLTSGPSGPGDGKTPPGPAGQDGRGPPGPPGARGOAGVMGFPGP	333
378	Db	GLTSGPSGPGDGKTPPGPAGQDGRGPPGPPGARGOAGVMGFPGP	437
334	QY	GVPPPGVAGVPAGKDXGAGAGQPPGAPGAGERGEOGPAGSPGFQGLPGPAGPPGBAGKP	393
438	Db	GVPPPGVAGVPAGKDXGAGAGQPPGAPGAGERGEOGPAGSPGFQGLPGPAGPPGBAGKP	497
394	QY	GEQGVPGDLGAPGSPGPAE-----PGP-----TGLPGPP	423
498	Db	GEQGVPGDLGAPGSPGARGERFFGGRGVQCPGAPGPRGANGAPGNDGAKGDAGAPGAP	557
424	QY	GERGPGGRGPPGADVAGPKPPAGRGSGPCGAPGKSGPCE-----AGRPGBAGLP	474
558	Db	GSQAGPLQGMHPPGERGAAGLPCPKGDRGDAGPKGDSPKDGVRLTGP	617
475	QY	GAK-----GLTSGPSGPPDGKTPPGPAGQDGRGPPGPPGARGO-----	515
618	Db	GDKGESGPGPAGPTGARGAPGDRGEFPGPAGFAGPADQOPGAKGEPGDAGAKGDA	677
516	QY	-----AGVMGPPKGAAGEGPKAGERGVPPGCA-----VGPAGKDXGAGAGPP	561
678	Db	GPDPGAPGAPPGPIGNVGAFGAKGARGSGAPPGATGFGCAAGRVGPPGPGSGNAGPPGPP	737
562	QY	GPAGP-----AGRGEOPAGSPGFQGLPGPAGPCEACKPCEQGVPGDLGAPGSPGAGBP	618
738	Db	GPAGKEGKGPRGETGPAGRCEGVPPGPPGAGEKSGPADGAPACATGTFGPGQTAGOR	797
619	QY	GPTGLPGPPGGRGPGSRGFPFGADVAGPKPPAGRGSGPCGAPG-----KSGPGBAGR	675
798	Db	GVVGLPGQGRGPPGLPGPSGEPKQKQPSGASGERGPPGPMGPPGLAGPPSGSGREGAP	857
676	QY	GLPGAKGLTSGPSGPPDGKTPPGPAGQDGRGPPGPPGARGOAGVMGFPGPKGAGBP	735
858	Db	GAEPSGPRGDSFGAKGDRGETGPAAGPAGPAGPAGVGPAGKSGDRGETGPAAGPV	917
736	QY	GKAGERGVGPPGPAVGPAKGDE-----AGACQPPGP-----AGPA	771
918	Db	GPAGARGAGPAGPQGRDKETGTEGQDGRGICKHRCFSLQCPGPPGSGPQSGSAGPA	977
772	QY	GERGSGGPGAGSP---GFQGLPGAPPGPGEAGKPGEOGVPGDLGAPGSPGAG	820
978	Db	GRGPPGSGAGAPGKQGLNGLPCPTGPPGPRGRTGDAGVPGPPGPPGPPGPPG	1029

RESULT 13

US-10-104-889-10
; Sequence 10, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS

BROKAW, JANE
ZHANG, GUANGHUI

PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD

Db	419	GPQPCGGPPGPKGNSGPFAGPSKGDTCAGKEGPGVGVQGPFGAGEGKRGARCEPPT	478
Qy	214	GLPCCPBERGCGPSGRGPPGADGVAGPKGPACGERSGPGAPGKSGPGEAGRPGEAGLPGAK	273
Db	479	GLPCCPBERGCGPSGRGPPGADGVAGPKGPACGERSGPGAPGKSGPGEAGRPGEAGLPGAK	538
Qy	274	GLTGSPPSGPDPGKTGTPPGPAGQDGRPGPPPGARGQAGVMGFPKPKGAAGEPKAGER	333
Db	539	GLTGSPPSGPDPGKTGTPPGPAGQDGRPGPPPGARGQAGVMGFPKPKGAAGEPKAGER	598
Qy	334	GVPPGPGAVGAGKDKBAGAGQGGPPGAPGAGERGEGGAGSGFGQLPGPPAGPPGEGAKP	393
Db	599	GVPPGPGAVGAGKDKBAGAGQGGPPGAPGAGERGEGGAGSGFGQLPGPPAGPPGEGAKP	658
Qy	394	GEQVPPGDLNAGPSGSGPAGE-----PGP-----TCLPAPP	423
Db	659	GEQVPPGDLNAGPSGSGAGERGFPCEGVRGVQGGPPGAPRGANGAPNDNGAKGDAGAPAP	718
Qy	424	GERGGPSRGGPPGADGVAGPKGPACGERSGSPGAPKSGSGE-----AGRPGCEAGLP	474
Db	719	GSQAGPGLQGMGPRGGAAGLPKPKGDRDADGPKGADGSPGKDGVRGLTGPITGPPGPAGAP	778
Qy	475	GAK-----GLTGSPPSGPPDGKTPGPPAGQDGRPGPPGPPGARGQ-----	515
Db	779	GDKGESGSPGAPGTGARGAFGDRGEGPPGPPGAPGAGPPGADGQPCAKGEGPCDGA	838
Qy	516	-----AGVMGPPGPKGAAGEPKGAGERGVVPPGPGA-----VGPAGKDGEGAGAGPP	561
Db	839	GPPGPAGPAGPPGPIGNVGAFGAKGARGSGAPPGATGPPGAAGRVPGPSGNAGPPGPP	898
Qy	562	GPAGP-----AGERGCGPAGSGPFGQLPGPAGPPGEGAGKPGGCGVPDGLGAPGPSG	618
Db	899	GPAGKEGKGPRGEGTGPAGRFGEGVPPGPPPGPAGEKSGPAGDPAGAPGTPGPGIAGQR	958
Qy	619	GPTCLGPCCPGRGSGRGGFPAGDGVAGPKGPACGERSGSPGAPG-----KSGPCEAGR	675
Db	959	GVWGLPQORGBRGGFPGLPSPGSEPGKQGPSGASGERGPPGMPGLAGPPGSGSREGAP	1018
Qy	676	GLPKAKGLTSGPSGSPGPDGKTGTPPGPAGQDGRPGPPGARGQAGVMGFPKPKGAAGEP	735
Db	1019	GAEGSPRDSFGAKDRGETGTGAPGPCAPCAPGAPVGVGAKSGGDRGETGTGAPGAPV	1078
Qy	736	GKAGERGVPPGPGAVGPAKGDS-----AGAQGGPPG-----AGPA	771
Db	1079	GPACARGAPGPPGPRDKGTEGQCDRGIKGHRGFSGLQGPPGPPGSGPQEGSPGASGPA	1138
Qy	772	GERGEGQAGSP-----FGQLPGPAGPPGEGAKPGGEGQGVPGDLGAPGSGPAG	820
Db	1139	GPRGPPGSGAPGKDGKGLNGLPGPIGPPGPRGRTGDAGVWPPGPPGPPGPPGPPGPPGPP	1190

```

RESULT 15
US-10-357-851-1
; Sequence 1, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-1

```

Query Match

59.6%: Score 2765: DB 16: Length 1464:

Search completed: June 17, 2005, 15:35:47

Best Local Similarity 55.4%; Pred. No. 3.9e-131;					
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;					
QY	1	GPPGEPPTGLPDPGERGGPSRGPFGADGVAGPKGPAGERGSPGAPGKSGPGA---	57		
Db	179	GPWGPSGRGLPCPPCGA PGQFQGPGPEGPASGPMGRP GPCTGPKNGDGDGAGKP	238		
QY	58	GRPGEAGLP---GAKLGTGSPGSPGP-----DGKTGPPCPAQDQRPGPPGPGAR	105		
Db	239	GRPGERGPPCQCARGLPCTAGLPGMKHRRGSLGDAKGADGAPGAPGKGPSPGENGAP	298		
QY	106	GQAGVMGFPGPKGAACGCKACGER-----GVPGPAGVGA-----GKDGRAGA Q	150		
Db	299	GQMGPRLGPERGRCPGAPGAGNDGATGAAGPPPTGTPAGPPFP GAVGAKGAGAPQ	358		
QY	151	GPPGPAGPAGERGEOGF-----AGS PFGQLPGPA 180			
Db	359	GPRGSFGPVRG PCPPGPAAGPAGNPDGQPGAKGANGAPCIAGAPGPGARGPS	418		
QY	181	GPPEGAGKPGBOGVPODLGAP-----GPSGPA-----GSPGPT 213			
Db	419	GPQCGGPPGPKNSGEPAPSKTGTAKEGPPVGVOQPPGPAGEEKKRGARGEGPT	478		
QY	214	GLPGPCEGGRCSRCFGADGVAGPKGPAGBERGSPGAPGKSGCEAGRPGBAGLPGAK	273		
Db	479	GLFPGPERGGSRGFFCADGVAGPKGPAGBERGSPGAPGKSGCEAGRPGBAGLPGAK	538		
QY	274	GLTSGSPGPDCKTTPPPAGQDGRPGPPPGARGQAGVMGFPFKAAAGPPGKAGER	333		
Db	539	GLTSGSPGPDCKTCTPPGADGRPGPPPGARGQAGVMGFPFKAAAGPPGKAGER	598		
QY	334	GVPGPAGVGPAGKDEAGAQQP PPAGPAGERGEOGPAGSPGFQGLPGPAGPPGAGKP	393		
Db	599	GVPGPAGVGPAGKDEAGAOQPPGAPGAGERGEOGPAGSPGFQGLPGPAGPPGAGKP	658		
QY	394	GEOGVPDILGAPSPGAGE-----PGP-----TGLPGPP 423			
Db	659	GEOGVPDILGAPSPGARGERPPGBERGVQPPGPAGPRGANCA PNGDKAGDAGAPGAP	718		
QY	424	GERGFSRSGFFCADGVAGPKGPAGBERGSPGAPGKSGPGB-----AGRPGEAGLP	474		
Db	719	GSQGAFLQMPGECGAAGLPGPKDRGDAGPKGADGSPGKDGVRGLTGTFIPGPAGAP	778		
QY	475	GAK-----GLTSGSPGPDGKTGPPGPAGQDGRPGPPPPGARGO-----	515		
Db	779	GDKGESGSPGAPTGAARGPDGRGPPGPPGAFAGPFAGDQPGAKGPDAGAKGDA	838		
QY	516	-----AGVMFGPKCAACEBKAGERGVPPGPA-----VGPAGKDGAGAGQGP	561		
Db	839	GPPGPAGPAGPPGI GNVA GAPKAARGSAGPPGATGTFPGAARVPPGPPSNAGPPPPP	898		
QY	562	GPAGP---AGERGEOGPAGSPGQGLPGPAGPGEAGKPGEQCVPDGLAGPSPGAGEP	618		
Db	899	GPAGKEGKGPRGETGPAGRPGEVPPGPGPAGEKSGPCADGPAGACTGTPQGIAGOR	958		
QY	619	GPTGLPPOPERRGGPSRGFPGADGVAGPKGPAGERGSPGAPG---KSGPGBAGRPGEA	675		
Db	959	GVVGLPCQGERGFPLGPLSPSGPGKQSPSGASGERGPPGPMGPPGLAGPPGESGREGAP	1018		
QY	676	GLPGAKLGTGSPGSPCDGKTGPPGADGDRPGPPGPCARGQAQVMGPPGPKGAAGEP	735		
Db	1019	GABSGPRDGSFAKGDRETGPAGPPGAPGAPGAPVGPAGKSQDRGETGPAGPAGPV	1078		
QY	736	GKAGERGVPPGAVGPAFKDGB-----AGAQQPPG-----AGPA 771			
Db	1079	GPAGARGAPQPCRD KGETGECQDRIGIKHHRGFSLGQGP GPSPGCBQGPSGASGPA	1138		
QY	772	GERGEOGPAGSP---GFOGLPGPAGPPBAGKPGBOGVFDGLAGPSPGAP 820			
Db	1139	GPRGPPGSAAGPKDGLNLGPPGIGPGRGRTGDAGVGP GP GP GP GP GP GP GP	1190		

Search completed: June 17, 2005, 15:35:47

Fri Jun 17 15:44:49 2005

us-10-658-989a-3.rapb

Page 12

Job time : 102.932 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:42:55 ; Search time 28.1036 Seconds
(without alignments)
2810.812 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGPGTGLPGRGGRG.....GEOGVPGDLGAPGSPGAG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2756	59.4	1464	1 CGHUIS	collagen alpha 1(I)
2	2732.5	58.9	1453	1 S21626	collagen alpha 1(I)
3	2679	57.7	1042	1 CGCH1S	collagen alpha 1(I)
4	2477	53.4	1418	2 T45467	collagen alpha 1(I)
5	2476	53.4	1487	1 CGHU6C	collagen alpha 1(I)
6	2460	53.0	1419	2 A41182	collagen alpha 1(I)
7	2401	51.7	1486	1 B40333	collagen alpha 1(I)
8	2401	51.7	1486	1 B40333	collagen alpha 1(I)
9	2389	51.5	1492	2 A40333	collagen alpha 1(I)
10	2379	51.3	1049	1 CGBO7S	collagen alpha 1(I)
11	2377.5	51.2	1464	2 S9856	collagen alpha 1(I)
12	2370	51.1	1466	1 CGHU7L	collagen alpha 1(I)
13	2326	50.1	1496	1 CGHU2V	collagen alpha 1(I)
14	2303.5	49.6	1497	2 I49607	procollagen type V
15	2300	49.6	1373	1 A43291	collagen alpha 2(I)
16	2263.5	48.8	1366	1 CGHU2S	collagen alpha 2(I)
17	2235	48.2	671	1 CGRT1S	collagen alpha 1(I)
18	2229.5	48.0	1838	1 CGHU1V	collagen alpha 1(I)
19	2208.5	47.6	1843	2 S18803	collagen alpha 1(I)
20	2193.5	47.3	779	1 CGBO1S	collagen alpha 1(I)
21	2169	46.7	1806	1 CGHU1X	collagen alpha 1(I)
22	2161.5	46.6	886	2 I50694	collagen alpha 1(I)
23	2136.5	46.0	1414	1 S23809	collagen alpha 2(I)
24	2074.5	44.7	1027	2 S28774	collagen alpha cha
25	2071	44.6	2944	2 A54849	collagen alpha 1(I)
26	2060.5	44.4	1546	1 CGHU2E	collagen alpha 2(I)
27	2058	44.4	1691	1 S22917	collagen alpha 5(I)
28	2046	44.1	1690	1 CGHU1B	collagen alpha 4(I)
29	2036	43.9	1549	2 I48103	type VII collagen

30 2030 43.8 1763 2 S16366 collagen alpha 2(I)
31 1996 43.0 1669 1 CGHU4B collagen alpha 1(I)
32 1985 42.8 1669 1 CGMS4B collagen alpha 1(I)
33 1948.5 42.0 1670 1 CGHU3B collagen alpha 3(I)
34 1944.5 41.9 1603 2 S23810 collagen alpha 1(X)
35 1930.5 41.6 1712 1 CGHU2B collagen alpha 2(I)
36 1915.5 41.3 1759 2 T29351 collagen alpha 2(I)
37 1910 41.2 1758 2 T29350 hypothetical prote
38 1908.5 41.1 1752 2 A45407 collagen alpha 3(I)
39 1897 40.9 1024 2 S18251 collagen alpha 1(X)
40 1889 40.7 1707 2 A33526 collagen alpha 2(I)
41 1870.5 40.3 1775 2 A31893 collagen alpha 1(I)
42 1869 40.3 1744 2 S40991 collagen alpha 1(I)
43 1838.5 39.6 812 2 S31521 collagen COL1 - f
44 1833.5 39.5 1747 2 A54121 collagen alpha-4 c
45 1832 39.5 1691 1 CGHU6B collagen alpha 6(I)

ALIGNMENTS

CGHUIS
collagen alpha 1(I) chain precursor - human
N:Alternate names: procollagen alpha 1(I) chain
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text change 09-Jul-2004
C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11:
5269; A29439; I53466; A02852; I37247
R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988
A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex
A:Reference number: I60114; MUID:88329734; PMID:2843432
A:Accession: I60114
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369, 'L', 371-589 <DAL>
A:Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP
R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock
Biochem. J. 253, 919-922, 1988
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t
A:Reference number: S01143; MUID:89025644; PMID:3178743
A:Accession: S01143
A:Molecule type: mRNA
A:Residues: 1-472 <TRO>
A:Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NID:
A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;
Nature 310, 337-340, 1984
A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A:Reference number: A93335; MUID:84270697; PMID:6462220
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58, 'Q', 60-181 <CHU>
A:Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA25394.1; PID:G35658
R:Rossouw, C.M.S.; Vergeer, W.P.; du Flooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.
J. Biol. Chem. 263, 15151-15157, 1987
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enha
A:Reference number: I55254; MUID:88033098; PMID:2822714
A:Accession: I55254
A>Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control of
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
 A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter and enhancer regions identified.
 A:Reference number: 155237; MUID:85130970; PMID:2857713
 A:Accession: 155237
 A>Status: translation not shown; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-34 <CH2>
 A:Cross-references: GB:M10627; NID:G180383; PIDN:AAA51992.1; PID:G553226
 R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, J. Biol. Chem. 265, 6312-6317, 1990
 A>Title: In vivo and in vitro noncovalent association of excised alpha 1 (I) amino-terminal propeptides.
 A:Reference number: A35233; MUID:90202908; PMID:2318855
 A:Accession: A35233
 A:Molecule type: protein
 A:Residues: 33-52 <WR>
 A>Note: This propeptide fragment remained non-covalently bound to a defective, uncleaved propeptide.
 R:Well, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F. EMBO J. 8, 1705-1710, 1989
 A>Title: A base substitution in the exon of a collagen gene causes alternative splicing of the alpha 1(I) chain.
 A:Reference number: S09400; MUID:89356643; PMID:2767050
 A:Accession: S09400
 A:Molecule type: mRNA
 A:Residues: 156-183 <WEI>
 R:Click, E.M.; Bornstein, P. Biochemistry 9, 4699-4706, 1970
 A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) chain of collagen.
 A:Reference number: A90567; MUID:71038625; PMID:5529814
 A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
 A:Accession: B90567
 A:Molecule type: protein
 A:Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z', 235, 'Z', 237, 'Z', 239, 'Z', 241, 'Z', 243, 'Z', 245, 'Z', 247, 'Z', 249, 'Z', 251, 'Z', 253, 'Z', 255, 'Z', 257, 'Z', 259, 'Z', 261, 'Z', 263, 'Z', 265, 'Z', 267, 'Z', 269, 'Z', 271, 'Z', 273, 'Z', 275, 'Z', 277, 'Z', 279, 'Z', 281, 'Z', 283, 'Z', 285, 'Z', 287, 'Z', 289, 'Z', 291, 'Z', 293, 'Z', 295, 'Z', 297, 'Z', 299, 'Z', 301, 'Z', 303, 'Z', 305, 'Z', 307, 'Z', 309, 'Z', 311, 'Z', 313, 'Z', 315, 'Z', 317, 'Z', 319, 'Z', 321, 'Z', 323, 'Z', 325, 'Z', 327, 'Z', 329, 'Z', 331, 'Z', 333, 'Z', 335, 'Z', 337, 'Z', 339, 'Z', 341, 'Z', 343, 'Z', 345, 'Z', 347, 'Z', 349, 'Z', 351, 'Z', 353, 'Z', 355, 'Z', 357, 'Z', 359, 'Z', 361, 'Z', 363, 'Z', 365, 'Z', 367, 'Z', 369, 'Z', 371, 'Z', 373, 'Z', 375, 'Z', 377, 'Z', 379, 'Z', 381, 'Z', 383, 'Z', 385, 'Z', 387, 'Z', 389, 'Z', 391, 'Z', 393, 'Z', 395, 'Z', 397, 'Z', 399, 'Z', 401, 'Z', 403, 'Z', 405, 'Z', 407, 'Z', 409, 'Z', 411, 'Z', 413, 'Z', 415, 'Z', 417, 'Z', 419, 'Z', 421, 'Z', 423, 'Z', 425, 'Z', 427, 'Z', 429, 'Z', 431, 'Z', 433, 'Z', 435, 'Z', 437, 'Z', 439, 'Z', 441, 'Z', 443, 'Z', 445, 'Z', 447, 'Z', 449, 'Z', 451, 'Z', 453, 'Z', 455, 'Z', 457, 'Z', 459, 'Z', 461, 'Z', 463, 'Z', 465, 'Z', 467, 'Z', 469, 'Z', 471, 'Z', 473, 'Z', 475, 'Z', 477, 'Z', 479, 'Z', 481, 'Z', 483, 'Z', 485, 'Z', 487, 'Z', 489, 'Z', 491, 'Z', 493, 'Z', 495, 'Z', 497, 'Z', 499, 'Z', 501, 'Z', 503, 'Z', 505, 'Z', 507, 'Z', 509, 'Z', 511, 'Z', 513, 'Z', 515, 'Z', 517, 'Z', 519, 'Z', 521, 'Z', 523, 'Z', 525, 'Z', 527, 'Z', 529, 'Z', 531, 'Z', 533, 'Z', 535, 'Z', 537, 'Z', 539, 'Z', 541, 'Z', 543, 'Z', 545, 'Z', 547, 'Z', 549, 'Z', 551, 'Z', 553, 'Z', 555, 'Z', 557, 'Z', 559, 'Z', 561, 'Z', 563, 'Z', 565, 'Z', 567, 'Z', 569, 'Z', 571, 'Z', 573, 'Z', 575, 'Z', 577, 'Z', 579, 'Z', 581, 'Z', 583, 'Z', 585, 'Z', 587, 'Z', 589, 'Z', 591, 'Z', 593, 'Z', 595, 'Z', 597, 'Z', 599, 'Z', 601, 'Z', 603, 'Z', 605, 'Z', 607, 'Z', 609, 'Z', 611, 'Z', 613, 'Z', 615, 'Z', 617, 'Z', 619, 'Z', 621, 'Z', 623, 'Z', 625, 'Z', 627, 'Z', 629, 'Z', 631, 'Z', 633, 'Z', 635, 'Z', 637, 'Z', 639, 'Z', 641, 'Z', 643, 'Z', 645, 'Z', 647, 'Z', 649, 'Z', 651, 'Z', 653, 'Z', 655, 'Z', 657, 'Z', 659, 'Z', 661, 'Z', 663, 'Z', 665, 'Z', 667, 'Z', 669, 'Z', 671, 'Z', 673, 'Z', 675, 'Z', 677, 'Z', 679, 'Z', 681, 'Z', 683, 'Z', 685, 'Z', 687, 'Z', 689, 'Z', 691, 'Z', 693, 'Z', 695, 'Z', 697, 'Z', 699, 'Z', 701, 'Z', 703, 'Z', 705, 'Z', 707, 'Z', 709, 'Z', 711, 'Z', 713, 'Z', 715, 'Z', 717, 'Z', 719, 'Z', 721, 'Z', 723, 'Z', 725, 'Z', 727, 'Z', 729, 'Z', 731, 'Z', 733, 'Z', 735, 'Z', 737, 'Z', 739, 'Z', 741, 'Z', 743, 'Z', 745, 'Z', 747, 'Z', 749, 'Z', 751, 'Z', 753, 'Z', 755, 'Z', 757, 'Z', 759, 'Z', 761, 'Z', 763, 'Z', 765, 'Z', 767, 'Z', 769, 'Z', 771, 'Z', 773, 'Z', 775, 'Z', 777, 'Z', 779, 'Z', 781, 'Z', 783, 'Z', 785, 'Z', 787, 'Z', 789, 'Z', 791, 'Z', 793, 'Z', 795, 'Z', 797, 'Z', 799, 'Z', 801, 'Z', 803, 'Z', 805, 'Z', 807, 'Z', 809, 'Z', 811, 'Z', 813, 'Z', 815, 'Z', 817, 'Z', 819, 'Z', 821, 'Z', 823, 'Z', 825, 'Z', 827, 'Z', 829, 'Z', 831, 'Z', 833, 'Z', 835, 'Z', 837, 'Z', 839, 'Z', 841, 'Z', 843, 'Z', 845, 'Z', 847, 'Z', 849, 'Z', 851, 'Z', 853, 'Z', 855, 'Z', 857, 'Z', 859, 'Z', 861, 'Z', 863, 'Z', 865, 'Z', 867, 'Z', 869, 'Z', 871, 'Z', 873, 'Z', 875, 'Z', 877, 'Z', 879, 'Z', 881, 'Z', 883, 'Z', 885, 'Z', 887, 'Z', 889, 'Z', 891, 'Z', 893, 'Z', 895, 'Z', 897, 'Z', 899, 'Z', 901, 'Z', 903, 'Z', 905, 'Z', 907, 'Z', 909, 'Z', 911, 'Z', 913, 'Z', 915, 'Z', 917, 'Z', 919, 'Z', 921, 'Z', 923, 'Z', 925, 'Z', 927, 'Z', 929, 'Z', 931, 'Z', 933, 'Z', 935, 'Z', 937, 'Z', 939, 'Z', 941, 'Z', 943, 'Z', 945, 'Z', 947, 'Z', 949, 'Z', 951, 'Z', 953, 'Z', 955, 'Z', 957, 'Z', 959, 'Z', 961, 'Z', 963, 'Z', 965, 'Z', 967, 'Z', 969, 'Z', 971, 'Z', 973, 'Z', 975, 'Z', 977, 'Z', 979, 'Z', 981, 'Z', 983, 'Z', 985, 'Z', 987, 'Z', 989, 'Z', 991, 'Z', 993, 'Z', 995, 'Z', 997, 'Z', 999, 'Z', 1001, 'Z', 1003, 'Z', 1005, 'Z', 1007, 'Z', 1009, 'Z', 1011, 'Z', 1013, 'Z', 1015, 'Z', 1017, 'Z', 1019, 'Z', 1021, 'Z', 1023, 'Z', 1025, 'Z', 1027, 'Z', 1029, 'Z', 1031, 'Z', 1033, 'Z', 1035, 'Z', 1037, 'Z', 1039, 'Z', 1041, 'Z', 1043, 'Z', 1045, 'Z', 1047, 'Z', 1049, 'Z', 1051, 'Z', 1053, 'Z', 1055, 'Z', 1057, 'Z', 1059, 'Z', 1061, 'Z', 1063, 'Z', 1065, 'Z', 1067, 'Z', 1069, 'Z', 1071, 'Z', 1073, 'Z', 1075, 'Z', 1077, 'Z', 1079, 'Z', 1081, 'Z', 1083, 'Z', 1085, 'Z', 1087, 'Z', 1089, 'Z', 1091, 'Z', 1093, 'Z', 1095, 'Z', 1097, 'Z', 1099, 'Z', 1101, 'Z', 1103, 'Z', 1105, 'Z', 1107, 'Z', 1109, 'Z', 1111, 'Z', 1113, 'Z', 1115, 'Z', 1117, 'Z', 1119, 'Z', 1121, 'Z', 1123, 'Z', 1125, 'Z', 1127, 'Z', 1129, 'Z', 1131, 'Z', 1133, 'Z', 1135, 'Z', 1137, 'Z', 1139, 'Z', 1141, 'Z', 1143, 'Z', 1145, 'Z', 1147, 'Z', 1149, 'Z', 1151, 'Z', 1153, 'Z', 1155, 'Z', 1157, 'Z', 1159, 'Z', 1161, 'Z', 1163, 'Z', 1165, 'Z', 1167, 'Z', 1169, 'Z', 1171, 'Z', 1173, 'Z', 1175, 'Z', 1177, 'Z', 1179, 'Z', 1181, 'Z', 1183, 'Z', 1185, 'Z', 1187, 'Z', 1189, 'Z', 1191, 'Z', 1193, 'Z', 1195, 'Z', 1197, 'Z', 1199, 'Z', 1201, 'Z', 1203, 'Z', 1205, 'Z', 1207, 'Z', 1209, 'Z', 1211, 'Z', 1213, 'Z', 1215, 'Z', 1217, 'Z', 1219, 'Z', 1221, 'Z', 1223, 'Z', 1225, 'Z', 1227, 'Z', 1229, 'Z', 1231, 'Z', 1233, 'Z', 1235, 'Z', 1237, 'Z', 1239, 'Z', 1241, 'Z', 1243, 'Z', 1245, 'Z', 1247, 'Z', 1249, 'Z', 1251, 'Z', 1253, 'Z', 1255, 'Z', 1257, 'Z', 1259, 'Z', 1261, 'Z', 1263, 'Z', 1265, 'Z', 1267, 'Z', 1269, 'Z', 1271, 'Z', 1273, 'Z', 1275, 'Z', 1277, 'Z', 1279, 'Z', 1281, 'Z', 1283, 'Z', 1285, 'Z', 1287, 'Z', 1289, 'Z', 1291, 'Z', 1293, 'Z', 1295, 'Z', 1297, 'Z', 1299, 'Z', 1301, 'Z', 1303, 'Z', 1305, 'Z', 1307, 'Z', 1309, 'Z', 1311, 'Z', 1313, 'Z', 1315, 'Z', 1317, 'Z', 1319, 'Z', 1321, 'Z', 1323, 'Z', 1325, 'Z', 1327, 'Z', 1329, 'Z', 1331, 'Z', 1333, 'Z', 1335, 'Z', 1337, 'Z', 1339, 'Z', 1341, 'Z', 1343, 'Z', 1345, 'Z', 1347, 'Z', 1349, 'Z', 1351, 'Z', 1353, 'Z', 1355, 'Z', 1357, 'Z', 1359, 'Z', 1361, 'Z', 1363, 'Z', 1365, 'Z', 1367, 'Z', 1369, 'Z', 1371, 'Z', 1373, 'Z', 1375, 'Z', 1377, 'Z', 1379, 'Z', 1381, 'Z', 1383, 'Z', 1385, 'Z', 1387, 'Z', 1389, 'Z', 1391, 'Z', 1393, 'Z', 1395, 'Z', 1397, 'Z', 1399, 'Z', 1401, 'Z', 1403, 'Z', 1405, 'Z', 1407, 'Z', 1409, 'Z', 1411, 'Z', 1413, 'Z', 1415, 'Z', 1417, 'Z', 1419, 'Z', 1421, 'Z', 1423, 'Z', 1425, 'Z', 1427, 'Z', 1429, 'Z', 1431, 'Z', 1433, 'Z', 1435, 'Z', 1437, 'Z', 1439, 'Z', 1441, 'Z', 1443, 'Z', 1445, 'Z', 1447, 'Z', 1449, 'Z', 1451, 'Z', 1453, 'Z', 1455, 'Z', 1457, 'Z', 1459, 'Z', 1461, 'Z', 1463, 'Z', 1465, 'Z', 1467, 'Z', 1469, 'Z', 1471, 'Z', 1473, 'Z', 1475, 'Z', 1477, 'Z', 1479, 'Z', 1481, 'Z', 1483, 'Z', 1485, 'Z', 1487, 'Z', 1489, 'Z', 1491, 'Z', 1493, 'Z', 1495, 'Z', 1497, 'Z', 1499, 'Z', 1501, 'Z', 1503, 'Z', 1505, 'Z', 1507, 'Z', 1509, 'Z', 1511, 'Z', 1513, 'Z', 1515, 'Z', 1517, 'Z', 1519, 'Z', 1521, 'Z', 1523, 'Z', 1525, 'Z', 1527, 'Z', 1529, 'Z', 1531, 'Z', 1533, 'Z', 1535, 'Z', 1537, 'Z', 1539, 'Z', 1541, 'Z', 1543, 'Z', 1545, 'Z', 1547, 'Z', 1549, 'Z', 1551, 'Z', 1553, 'Z', 1555, 'Z', 1557, 'Z', 1559, 'Z', 1561, 'Z', 1563, 'Z', 1565, 'Z', 1567, 'Z', 1569, 'Z', 1571, 'Z', 1573, 'Z', 1575, 'Z', 1577, 'Z', 1579, 'Z', 1581, 'Z', 1583, 'Z', 1585, 'Z', 1587, 'Z', 1589, 'Z', 1591, 'Z', 1593, 'Z', 1595, 'Z', 1597, 'Z', 1599, 'Z', 1601, 'Z', 1603, 'Z', 1605, 'Z', 1607, 'Z', 1609, 'Z', 1611, 'Z', 1613, 'Z', 1615, 'Z', 1617, 'Z', 1619, 'Z', 1621, 'Z', 1623, 'Z', 1625, 'Z', 1627, 'Z', 1629, 'Z', 1631, 'Z', 1633, 'Z', 1635, 'Z', 1637, 'Z', 1639, 'Z', 1641, 'Z', 1643, 'Z', 1645, 'Z', 1647, 'Z', 1649, 'Z', 1651, 'Z', 1653, 'Z', 1655, 'Z', 1657, 'Z', 1659, 'Z', 1661, 'Z', 1663, 'Z', 1665, 'Z', 1667, 'Z', 1669, 'Z', 1671, 'Z', 1673, 'Z', 1675, 'Z', 1677, 'Z', 1679, 'Z', 1681, 'Z', 1683, 'Z', 1685, 'Z', 1687, 'Z', 1689, 'Z', 1691, 'Z', 1693, 'Z', 1695, 'Z', 1697, 'Z', 1699, 'Z', 1701, 'Z', 1703, 'Z', 1705, 'Z', 1707, 'Z', 1709, 'Z', 1711, 'Z', 1713, 'Z', 1715, 'Z', 1717, 'Z', 1719, 'Z', 1721, 'Z', 1723, 'Z', 1725, 'Z', 1727, 'Z', 1729, 'Z', 1731, 'Z', 1733, 'Z', 1735, 'Z', 1737, 'Z', 1739, 'Z', 1741, 'Z', 1743, 'Z', 1745, 'Z', 1747, 'Z', 1749, 'Z', 1751, 'Z', 1753, 'Z', 1755, 'Z', 1757, 'Z', 1759, 'Z', 1761, 'Z', 1763, 'Z', 1765, 'Z', 1767, 'Z', 1769, 'Z', 1771, 'Z', 1773, 'Z', 1775, 'Z', 1777, 'Z', 1779, 'Z', 1781, 'Z', 1783, 'Z', 1785, 'Z', 1787, 'Z', 1789, 'Z', 1791, 'Z', 1793, 'Z', 1795, 'Z', 1797, 'Z', 1799, 'Z', 1801, 'Z', 1803, 'Z', 1805, 'Z', 1807, 'Z', 1809, 'Z', 1811, 'Z', 1813, 'Z', 1815, 'Z', 1817, 'Z', 1819, 'Z', 1821, 'Z', 1823, 'Z', 1825, 'Z', 1827, 'Z', 1829, 'Z', 1831, 'Z', 1833, 'Z', 1835, 'Z', 1837, 'Z', 1839, 'Z', 1841, 'Z', 1843, 'Z', 1845, 'Z', 1847, 'Z', 1849, 'Z', 1851, 'Z', 1853, 'Z', 1855, 'Z', 1857, 'Z', 1859, 'Z', 1861, 'Z', 1863, 'Z', 1865, 'Z', 1867, 'Z', 1869, 'Z', 1871, 'Z', 1873, 'Z', 1875, 'Z', 1877, 'Z', 1879, 'Z', 1881, 'Z', 1883, 'Z', 1885, 'Z', 1887, 'Z', 1889, 'Z', 1891, 'Z', 1893, 'Z', 1895, 'Z', 1897, 'Z', 1899, 'Z', 1901, 'Z', 1903, 'Z', 1905, 'Z', 1907, 'Z', 1909, 'Z', 1911, 'Z', 1913, 'Z', 1915, 'Z', 1917, 'Z', 1919, 'Z', 1921, 'Z', 1923, 'Z', 1925, 'Z', 1927, 'Z', 1929, 'Z', 1931, 'Z', 1933, 'Z', 1935, 'Z', 1937, 'Z', 1939, 'Z', 1941, 'Z', 1943, 'Z', 1945, 'Z', 1947, 'Z', 1949, 'Z', 1951, 'Z', 1953, 'Z', 1955, 'Z', 1957, 'Z', 1959, 'Z', 1961, 'Z', 1963, 'Z', 1965, 'Z', 1967, 'Z', 1969, 'Z', 1971, 'Z', 1973, 'Z', 1975, 'Z', 1977, 'Z', 1979, 'Z', 1981, 'Z', 1983, 'Z', 1985, 'Z', 1987, 'Z', 1989, 'Z', 1991, 'Z', 1993, 'Z', 1995, 'Z', 1997, 'Z', 1999, 'Z', 2001, 'Z', 2003, 'Z', 2005, 'Z', 2007, 'Z', 2009, 'Z', 2011, 'Z', 2013, 'Z', 2015, 'Z', 2017, 'Z', 2019, 'Z', 2021, 'Z', 2023, 'Z', 2025, 'Z', 2027, 'Z', 2029, 'Z', 2031, 'Z', 2033, 'Z', 2035, 'Z', 2037, 'Z', 2039, 'Z', 2041, 'Z', 2043, 'Z', 2045, 'Z', 2047, 'Z', 2049, 'Z', 2051, 'Z', 2053, 'Z', 2055, 'Z', 2057, 'Z', 2059, 'Z', 2061, 'Z', 2063, 'Z', 2065, 'Z', 2067, 'Z', 2069, 'Z', 2071, 'Z', 2073, 'Z', 2075, 'Z', 2077, 'Z', 2079, 'Z', 2081, 'Z', 2083, 'Z', 2085, 'Z', 2087, 'Z', 2089, 'Z', 2091, 'Z', 2093, 'Z', 2095, 'Z', 2097, 'Z', 2099, 'Z', 2101, 'Z', 2103, 'Z', 2105, 'Z', 2107, 'Z', 2109, 'Z', 2111, 'Z', 2113, 'Z', 2115, 'Z', 2117, 'Z', 2119, 'Z', 2121, 'Z', 2123, 'Z', 2125, 'Z', 2127, 'Z', 2129, 'Z', 2131, 'Z', 2133, 'Z', 2135, 'Z', 2137, 'Z', 2139, 'Z', 2141, 'Z', 2143, 'Z', 2145, 'Z', 2147, 'Z', 2149, 'Z', 2151, 'Z', 2153, 'Z', 2155, 'Z', 2157, 'Z', 2159, 'Z', 2161, 'Z', 2163, 'Z', 2165, 'Z', 2167, 'Z', 2169, 'Z', 2171, 'Z', 2173, 'Z', 2175, 'Z', 2177, 'Z', 2179, 'Z', 2181, 'Z', 2183, 'Z', 2185, 'Z', 2187, 'Z', 2189, 'Z', 2191, 'Z', 2193, 'Z', 2195, 'Z', 2197, 'Z', 2199, 'Z', 2201, 'Z', 2203, 'Z', 2205, 'Z', 2207, 'Z', 2209, 'Z', 2211, 'Z', 2213, 'Z', 2215, 'Z', 2217, 'Z', 2219, 'Z', 2221, 'Z', 2223, 'Z', 2225, 'Z', 2227, 'Z', 2229, 'Z', 2231, 'Z', 2233, 'Z', 2235, 'Z', 2237, 'Z', 2239, 'Z', 2241, 'Z', 2243, 'Z', 2245, 'Z', 2247, 'Z', 2249, 'Z', 2251, 'Z', 2253, 'Z', 2255, 'Z', 2257, 'Z', 2259, 'Z', 2261, 'Z', 2263, 'Z', 2265, 'Z', 2267, 'Z', 2269, 'Z', 2271, 'Z', 2273, 'Z', 2275, 'Z', 2277, 'Z', 2279, 'Z', 2281, 'Z', 2283, 'Z', 2285, 'Z', 2287, 'Z', 2289, 'Z', 2291, 'Z', 2293, 'Z', 2295, 'Z', 2297, 'Z', 2299, 'Z', 2301, 'Z', 2303, 'Z', 2305, 'Z', 2307, 'Z', 2309, 'Z', 2311, 'Z', 2313, 'Z', 2315, 'Z', 2317, 'Z', 2319, 'Z', 2321, 'Z', 2323, 'Z', 2325, 'Z', 2327, 'Z', 2329, 'Z', 2331, 'Z', 2333, 'Z', 2335, 'Z', 2337, 'Z', 2339, 'Z', 2341, 'Z', 2343, 'Z', 2345, 'Z', 2347, 'Z', 2349, 'Z', 2351, 'Z', 2353, 'Z', 2355, 'Z', 2357, 'Z', 2359, 'Z', 2361, 'Z', 2363, 'Z', 2365, 'Z', 2367, 'Z', 2369, 'Z', 2371, 'Z', 2373, 'Z', 2375, 'Z', 2377, 'Z', 2379, 'Z', 2381, 'Z', 2383, 'Z', 2385, 'Z', 2387, 'Z', 2389, 'Z', 2391, 'Z', 2393, 'Z', 2395, 'Z', 2397, 'Z', 2399, 'Z', 2401, 'Z', 2403, 'Z', 2405, 'Z', 2407, 'Z', 2409, 'Z', 2411, 'Z', 2413, 'Z', 2415, 'Z', 2417, 'Z', 2419, 'Z', 2421, 'Z', 2423, 'Z', 2425, 'Z', 2427, 'Z', 2429, 'Z', 2431, 'Z', 2433, 'Z', 2435, 'Z', 2437, 'Z', 2439, 'Z', 2441, 'Z', 2443, 'Z', 2445, 'Z', 2447, 'Z', 2449, 'Z', 2451, 'Z', 2453, 'Z', 2455, 'Z', 2457, 'Z', 2459, 'Z', 2461, 'Z', 2463, 'Z', 2465, 'Z', 2467, 'Z', 2469, 'Z', 2471, 'Z', 2473, 'Z', 2475, 'Z', 2477, 'Z', 2479, 'Z', 2481, 'Z', 2483, 'Z', 2485, 'Z', 2487, 'Z', 2489, 'Z', 2491, 'Z', 2493, 'Z', 2495, 'Z', 2497, 'Z', 2499, 'Z', 2501, 'Z', 2503, 'Z', 2505, 'Z', 2507, 'Z', 2509, 'Z', 2511, 'Z', 2513, 'Z', 2515, 'Z', 2517, 'Z', 2519, 'Z', 2521, 'Z', 2523, 'Z', 2525, 'Z', 2527, 'Z', 2529, 'Z', 2531, 'Z', 2533, 'Z', 2535, 'Z', 2537, 'Z', 2539, 'Z', 2541, 'Z', 2543, 'Z', 2545, 'Z', 2547, 'Z', 2549, 'Z', 2551, 'Z', 2553, 'Z', 2555, 'Z', 2557, 'Z', 2559, 'Z', 2561, 'Z', 2563, 'Z', 2565, 'Z', 2567, 'Z', 2569, 'Z', 2571, 'Z', 2573, 'Z', 2575, 'Z', 2577, 'Z', 2579, 'Z', 2581, 'Z', 2583, 'Z', 2585, 'Z', 2587, 'Z', 2589, 'Z', 2591, 'Z', 2593, 'Z', 2595, 'Z', 2597, 'Z', 2599, 'Z', 2601, 'Z', 2603, 'Z', 2605, 'Z', 2607, 'Z', 2609, 'Z', 2611, 'Z', 2613, 'Z', 2615, 'Z', 2617, 'Z', 2619, 'Z', 2621, 'Z', 2623, 'Z', 2625, 'Z', 2627, 'Z', 2629, 'Z', 2631, 'Z', 2633, 'Z', 2635, 'Z', 2637, 'Z', 2639, 'Z', 2641, 'Z', 2643, 'Z', 2645, 'Z', 2647, 'Z', 2649, 'Z', 2651, 'Z', 2653, 'Z', 2655, 'Z', 2657, 'Z', 2659, 'Z', 2661, 'Z', 2663, 'Z', 2665, 'Z', 2667, 'Z', 2669, 'Z', 2671, 'Z', 2673, 'Z', 2675, 'Z', 2677, 'Z', 2679, 'Z', 2681, 'Z', 2683, 'Z', 2685, 'Z', 2687, 'Z', 2689, 'Z', 2691, 'Z', 2693, 'Z', 2695, 'Z', 2697, 'Z', 2699, 'Z', 2701, 'Z', 2703, 'Z', 2705, 'Z', 2707, 'Z', 2709, 'Z', 2711, 'Z', 2713, 'Z', 2715, 'Z', 2717, 'Z', 2719, 'Z', 2721, 'Z', 2723, 'Z', 2725, 'Z', 2727, 'Z', 2729, 'Z', 2731, 'Z', 2733, 'Z', 2735, 'Z', 2737, 'Z', 2739, 'Z', 2741, 'Z', 2743, 'Z', 2745, 'Z', 2747, 'Z', 2749, 'Z', 2751, 'Z', 2753, 'Z', 2755, 'Z', 2757, 'Z', 2759, 'Z', 2761, 'Z', 2763, 'Z', 2765, 'Z', 2767, 'Z', 2769, 'Z', 2771, 'Z', 2773, 'Z', 2775, 'Z', 2777, 'Z', 2779, 'Z', 2781, 'Z', 2783, 'Z', 2785, 'Z', 2787, 'Z', 2789, 'Z', 2791, 'Z', 2793, 'Z', 2795, 'Z', 2797, 'Z', 2799, 'Z', 2801, 'Z', 2803, 'Z', 2805, 'Z', 2807, 'Z', 2809, 'Z', 2811, 'Z', 2813, 'Z', 2815, 'Z', 2817, 'Z', 2819, 'Z', 2821, 'Z', 2823, 'Z', 2825, 'Z', 2827, 'Z', 2829, 'Z', 2831, 'Z', 2833, 'Z', 2835, 'Z', 2837, 'Z', 2839, 'Z', 2841, 'Z', 2843, 'Z', 2845, 'Z', 2847, 'Z',

736 GXAGRGVGPAGVPAGKDG-----AGAQGGPGP-----AGPA 771
||| ||| ||| ||| ||| : |||
Db 1079 GPAGARGPAGPGRGDKGRTGQDRGIKGHRGFGSLQGPPGPSGEQGFSGASGPA 1138
||| ||| ||| ||| ||| : |||

772 GERGEGCGAGSP---GFGGLPGAPGPEAGKPGEQGVPGDLGAPGSPGAG 820
||| ||| ||| ||| ||| : |||
Db 1139 GPRPPGSAGAKOGKGLNLPGLPGPRGTGDAGPVGPPGPPGPPG 1190

RESULT 2

S21626 collagen alpha 1(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text_change 09-Jul-2004
A:Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R:L.I., S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A>Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I collagen
A:Reference number: S57243
A:Accession: S57243
A:Molecule type: mRNA
A:Residues: 1-1453 <LIS>
A:A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:g470673; PIDN:AAA88912.1; PID:g470673
R:Metzgerant, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A>Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16374
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MET>
A:A:Cross-references: EMBL:X57981; NID:g50484; PIDN:CAA41046.1; PID:g50485
R:French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A>Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A:Reference number: A23982; MUID:86137403; PMID:3841523
A:Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <PRE>
A:A:Cross-references: GB:M14423; NID:g192261; PIDN:AAA37333.1; PID:g192262
R:Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A>Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for a first intron of retrovirus into the first intron of alpha1(I) collagen gene leads to a new exon.
A:Reference number: I49559; MUID:83141374; PMID:6298597
A:Accession: I49559
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:A:Cross-references: GB:M17491; NID:g192263; PIDN:AAA37334.1; PID:g192264
R:Harbers, K.; Kuehn, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A>Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads to a new exon.
A:Reference number: I49557; MUID:84170331; PMID:6324198
A:Accession: I49557
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <RE2>
A:A:Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881
R:Fenton, S.P.; Laman, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8268229
A:Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>
A:A:Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881
R:Fenton, S.P.; Laman, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8268229
A:Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>
A:A:Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881
R:Fenton, S.P.; Laman, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: I48300; MUID:94344105; PMID:8065328
A:Accession: I48300
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

A:Cross-references: EMBL:X54876; NID:950486; PIDN:CAA38657.1; PID:950487
 C:Genetics:
 A:Gene: COL1A1
 A:introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>
 F:30-89/Domain: von Willebrand factor type C repeat homology <VMC>
 F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
 F:1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 58.9%; Score 2732.5; DB 2; Length 1453;
 Best Local Similarity 54.7%; Pred. No. 5.1e-129;
 Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;

Qy	1	GPPEGPTGLPPGE-----RGGPSRGFPFGADGVAGPKGPAERGSGPPA	48
Db	180	GPPGAPGQGFPGFPGEPGPGSGPMGPRGPPGPKNGDDGEAGKPRGPRGPPGQ	239
Qy	49	GPKSGPEAGR-----GEAGLPAKGLTSGPSGPDGKTGPPGAGQD	93
Db	240	GARLPGTAGLGMKXHRGSLDGAKGDAPGAPKGPSPGSGAPGQMGRLPGER	299
Qy	94	GRPGPPGPGARGQAVGMFGPKGAAGE-----PGKAGERGVPGPAGVAGPKDGEA	147
Db	300	GRPGPPGTAGARGNDGAVGAAGPGPTGTPGCFPFANVAKGAEAGQARGSEGGVVR	359
Qy	148	GAQPPPGPAGPAGRGQPP-----AGSPQFQLPGPAGPPGPAKPGQE	192
Db	360	GEPFPFGAAGPAGNPGADGQFGAKGANGAPGIAGAPFGARGPSGQPSGPPGPK	419
Qy	193	GVPGDLGAP-----GPSGA-----GEGPTGLPGPPRGGRGP	225
Db	420	NSGEPAPGNKGTGAKGPGATGVQGPFPAGEEGKRGARGEPSGLFPGPPSRGGP	479
Qy	226	GSRGFPDAGVAGKGPAGERSGPGAPKSGPCEAGRPGAEGLPAKGLTSGPSGPGD	285
Db	480	GSRGFPDAGVAGKGPAGERSGPGAPKSGPCEAGRPGAEGLPAKGLTSGPSGPD	539
Qy	286	GKTGPPGACDGRPGPPGARGQAVGMFGPKGACGAPKAGERGVPGPAGVGA	345
Db	540	GKTGPPGACDGRPGPPGARGQAVGMFGPKGTAGEPGKAGERGLPGPPGAVGA	599
Qy	346	GKDGEAGQPPGAPGAGERGQSPAGSPGFQGLPGPAGPCEAGKPGEQGVGDLGAP	405
Db	600	GKDGEAGQAPGAPGAGERGQSPAGSPGFQGLPGPAGPCEAGKPGEQGVGDLGAP	659
Qy	406	GPSGPAGE-----PGP-----TCLPDPGERGPGSRGPP	435
Db	660	GPSKAGERFPSERGVQGPFGPAGPRGNNGAPGNDGAKGDTGAPAGPSQAGPLQGMF	719
Qy	436	GADGVAGPKGPAGERSGPGAPKSGPGE-----AGRPGAEGLPAKGLTSGPSGP	486
Db	720	GERGAAGLPGKGDGRDAGFGKADGSPGDKGARGLTGPIGPPCPAGAPDKGAEAGSPGP	779
Qy	487	GPD-----CKTGPPGAGDGRPGPPPPGARGQ-----AGVMGPP	522
Db	780	GPTGARGAPGDRGEAGPPGAPGAPGADGDPGAKGEPDGTGVKDGADGPPGAPGAP	839
Qy	523	GPKGAAEPGKAGERGVPGPGA-----VGPAGKDGEAGACGPPCPAGP---	570
Db	840	GPIGNVGA GPKGPRAAGAPGATGPPGAAGRVGPPGPGSGNAGPPGPPGVKGGKGR	899
Qy	571	GEOGPAGSPGQGLPGAPGPGAEAGKPGQGVPIGAPGSPGAGEPFTGLPFPGER	630
Db	900	GETGPAGREGEVPPGPPGPGAEKSPGADGAPGSPGTGPGQIAGQGVVGLPQGER	959
Qy	631	GQPSGRFPDAGVAGPKGPAERSGPPGAP---KGSFGEAGRPGAEGLPAKGLTGP	687
Db	960	GFPLGPGSGEPKQKQPSGSSGERGPPGPMGPPGLAGPPGSGRSGSPGSGPRDGAP	1019
Qy	688	GSPPDGTGTPGPAQDGRPGPPGARGQAVGMFGFPKGAAGEPGKAGERGVPGP---	746


```
Db 383 GSPGPDGKTGPPGAGDGRGPGAGPFGARGQAGVMGFPFGKGAAGBPKGCGERAGCPGP 442
Qy 340 CAVGPAGKDGAGAGAGQPPGPPAGPAGERGEQGPAGSPGFGQLGPPAGPPGKAGKPGGEQGV 399
Db 443 CAVGAAGKDGAGAGAGQPPGPPGTPGAGERGEQGPAGAPFGQLGPPAGPPGKAGKPGGEQGV 502
Qy 400 GDLGARGSGPAGE-----PCP-----TGLPSPGERGPP 429
Db 503 GNAGAPGAGARGERGPFGERGVQPPGPPGPRGANGAPNDGAKGDAGAPGPNEGPP 562
Qy 430 GSRFPFGADGVAGPKGPAGERGSGPP-----AGPKGSGPAGERP-----GEA 471
Db 563 GLEWPFGERGAAGLPGAKGDRGDPGPKADGAPGKDLGLTGLTGPFGPAGAPGDKGEA 622
Qy 472 GLPGAAGLUTSGSPGSPGPKTGGPPGPA-----GDQGRGPP 507
Db 623 GPPGAGTGCARGAPDRGEPGPPGPAFGAGPGADGQPKAGKGETGDAGAKGDAGPPGPA 682
Qy 508 GPPGARGQAGVMGFPFGKGAAGEPKKAGRGVPPGPAVAGPKDKDGEAGAGPPGPA 567
Db 683 GPTGAPGAGZVGAPGPKGARGSGAGPPGATGPPGAAGRVGPPGSGNTGLPFGPPGPA 742
Qy 568 GE---RGQGPAGSPGFGQLGPPAGPPGKAGKPGGEQGVGDLGAPGPPGSPGAGEPPTGLP 624
Db 743 GSKGPRGETGACRGCEGEPGPPGPKGKSPGADGPIGAPTPGPGQIAGQGVVGLP 802
Qy 625 GPPCRRGPGSGRFGPADGVAGPKGPAGERGSPGAPG---KGSPEAGRPCEAGLPGAK 681
Db 803 GQGRGPFGLPGPSGEPKQGPSASCASGERGPPGPMGLAGPPGKAGREGAPGAEAP 862
Qy 682 GLTSGPSGPDGKTGPPGPAQDGRGPPGPPGARGQAGVMGFPFGKGAAGEPKKAGRG 741
Db 863 GRDGAAGPKGDRGETGAPGAPGAPGAPGVPAGKNGDRGETGAPGAPGPPGPA 922
Qy 742 GVPGPPGAVGPAKGDE-----AGAQGP-----GPAGPAGER 774
Db 923 GPAGQGRGDKGTGTEGQDRGMKHRGFGSLGQPPGPPGAPGBOGPPSGAGPAGPRGP 982
Qy 775 GEQGPAGSPGFGQLGPPAGPGEAGKPGGEQGVGDLGAPGSPGAP 820
Db 983 GSAGAAGKDLNGLGPIGPPGPRGRTGEVGVGPPGPPGPPG 1028

RESULT 4
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
R:Richardson, D.W.; Dodge, G.R.
A:Submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in eq
A:Reference number: 222977
A:Accession: T45467
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1418 <RIC>
A:Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:RAB05773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
```

```
Query Match 53.4%; Score 2477; DB 2; Length 1418;
Best Local Similarity 50.6%; Pred. No. 2.4e-116;
Matches 500; Conservative 48; Mismatches 272; Indels 168; Gaps 17;

Qy 1 GPPGPGTGLPGLPGERGSGRSGFPAGDGVAGPKGPAGERGSGDPAGPKGSPGAEAR 60
Db 156 GNPGEPPGVSGMPGPPGPPGPKGDDGAGPKGSGRGGPPGQCARGFPPTGLP 215
Qy 61 -----GEAGLPAKGLTQSP---GSPG-----DGKTPGPPA--- 90
Db 216 GVKHGRGVLGDAKGAAGAPGVKGSGSPGSGPMGPRGLPGERGRTGPAAGAR 275
```

```
Qy 91 GQDRPGRPPPGARGQAGVMGFPFGKGAAGEPKKAGRG-----GVPCP- 134
Db 276 GNDQGPAPGPPGVGPPAGGPGFPGAPGAKGEAGTARGPEGAQGRGEGTGTGSGCPGA 335
Qy 135 -----PCAVGPAGKDGAGAGAGQPPGPPGAPGAGERGEQGP-----AGSPFGQLP 177
Db 336 GAAGNPGTDTGIPGAKGSAGAPGIAGAPGFGPGRGPPGPGQATGFLGPKGQGTGFIAGFK 395
Qy 178 GPAGPPGAEAKPGBOQVPGDLGAPGSPGAGEPPTGLPDPGRRGPGSGRGGFFGADGVA 237
Db 396 GEQPKGEPGPPAGPQAGAPGAPGABEGKRGARGEPGAGVPVPGRGAAPGNRGGFFGQDGLA 455
Qy 238 GPKGPAGERSGSPGAPGKSGPGEAGRPDEAGLPCAKGLTGSPPSGPDGKTKGPPGAPGD 297
Db 456 GPKGAPGERGSPGLAGPKGANGDPGRFEPGLPARGLTGRPDAGPOGKVGPSPGAGED 515
Qy 298 GRPFPFPARQAGQAGVMGFPFGPKGAAGEPKKAGRGVPPGPAVAGPKDKDGEAGAGQPP 357
Db 516 GRPFPFPQAGRGQAGVMGFPFGPKGANGEPKAGKEKGLPGAPGLRGLPGDKGETGAAGFP 575
Qy 358 GPAGPAGERGEGPAGSPGFGQLGPPAGPPGAEAKPGEQGVGDLGAPGSPGAGE---P 414
Db 576 GPAGPAGERGEGQAGPSPGFGQLGPPGPPGEGGKPGDQGVPGGAGAPGLVPRGERGFP 635
Qy 415 GPTGLP-----GPPGERGPGSGRSGPPGADGVAGPKGA 447
Db 636 GERSGPAQLQAGARGLPGTPTDTPKASGAPGAPGAGQPPGLQGMPPGERGAAGIAGPK 695
Qy 448 GERSGPAGPKSGPGE-----AGRPGEAGLPGAKGLTGSPPSGPD 489
Db 596 GDRDGVKEGPEGAPGKDGGRGLTGTGPPGAPGANGKEGVEGPPGAGTAGARGAPGR 755
Qy 490 GKTGPPGAPGODGRGPPGPPGAPG-----GQAGVMGFPFGPKGAAGEP-----GKA 534
Db 756 GETGPPGAPGAPGADGDPGAKGEGQAGKGDAGAPGQPSGAPGQPTGVTGPK 815
Qy 535 GERGVGPPGA-----VGPAGKDEAGAGQPPGPA---GPAGERGEGPAGSPGQF 582
Db 816 GARGAQPPGATGTPGAAGRVGPPGSGNGNPFPFGSGKDGKPKGKAGSDSGPPGRAGDP 875
Qy 583 GLPAPGPPGKAGKPGBOQVPGDLGAPGSPGAGEPPTGLPDPGRRGPGSGRGGFFGAD 642
Db 876 GLQGPAGPPEKEGEPGDDGSPGDPGPPGQLAGQRLGVLPGQGRGERGFFGLGPPSGEP 935
Qy 643 GVAGPKGPAERSGSPGAPGKSGPCEAGRPCEAGLPGA-----KGLT 684
Db 936 GKQAGPAGSGDRGPPGVPGLTGPAGEPREGTTPGADGPPGRDGAAGVKGDRGEAGAL 995
Qy 685 GSPSGPDDGKTGPPGAPGD---GRGPPPPGARGQAGVMGFPFGKGAAGEPKAGE- 740
Db 996 GAPGAPPPGSPGAPGTPGKQGRGEAGAGQPMGPAAGARGLFGPPGQPRGDKGEAGEA 1055
Qy 741 -----RGVPPGPAVGPAGKDEAGAGQPPGAPGABERGEGPAGSPGFGQLGPPGA 792
Db 1056 GERGLKHRGFTGLQGLPGLPPEGSDQAGSAPGSGRPPGPPGVGSPGKDGANGIPGPI 1115
Qy 793 GPPGKAGKPGEQVPGDLGAPGSPGAPG 820
Db 1116 GPPGPRGRSGETGAPGPPGNGPPGPPG 1143

RESULT 5
CCHU6C
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen 1
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S631
7250; I37251; I37252; I37253; I37254; I55338; I59535; I61910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A:Title: The human type II procollagen gene: identification of an additional protein-codi
```


R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage, J. Eur. J. Biochem. 234, 125-131, 1995

A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil cartilage

A;Reference number: S63514; MUID:96096730; PMID:8529631

A;Accession: S63514

A;Molecule type: Protein

A;Residues: 243-261;575-590;756-763,'X',765-779 <PRA>

R;Tiller, G.E.; Weis, M.A.; Polumbo, P.A.; Gruber, H.E.; Rimooin, D.L.; Cohn, D.H.; Eyre, A.M. J. Hum. Genet. 56, 388-395, 1995

A;Title: An RNA-splicing mutation (G+5TVS20) in the type II collagen gene (COL2A1) in a family with osteogenesis imperfecta

A;Reference number: I38867; MUID:95150028; PMID:7847372

A;Accession: I38867

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TIL1>

A;Cross-references: EMBL:U15195; NID:9557053; PIDN:AAB60370.1; PID:9557054

R;Ramirez, P.

A;Reference number: S04892

A;Accession: S04892

A;Molecule type: mRNA

A;Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214 <RAM>

A;Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050

R;Viikula, M.; Peltonen, L.

FEBS Lett. 250, 171-174, 1989

A;Title: Structural analyses of the polymorphic area in type II collagen gene.

A;Reference number: S05000; MUID:93255561; PMID:2753125

A;Accession: S05000

A;Molecule type: DNA

A;Residues: 630-640,'A',642-785 <VIK2>

A;Cross-references: EMBL:X16159; NID:929951; PIDN:CAA34278.1; PID:91335018; PIDN:CAA34275.1

PIDN:CAA34283.1; PID:91335023; PIDN:CAA34284.1; PID:91335024

R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimooin, D.L.; Cohn, D.H.; Eyre, D.J. Biol. Chem. 267, 22522-22526, 1992

A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain protein of a patient with osteogenesis imperfecta

A;Reference number: A43309; MUID:93054548; PMID:1429602

A;Accession: A43309

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: DNA; mRNA

A;Residues: 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',1038-1052,'E',1054-1068,'T',']

A;Cross-references: GB:L00977; NID:9180812; PIDN:AAB3914.1; PID:9258774

A;Note: sequence extracted from NCBI backbone (NCBI:P117273); parts of this sequence were not included in the GenBank entry

A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis

R;Tiller, G.E.; Rimooin, D.L.; Murray, L.W.; Cohn, D.H.

Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990

A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual with osteogenesis imperfecta

A;Reference number: S16502; MUID:90251662; PMID:2339128

A;Accession: S16502

A;Molecule type: DNA

A;Residues: 1164-1184,'GPGSKDGANGIPQPI',1185-1199 <TIL2>

A;Cross-references: EMBL:M37126; NID:9180808; PIDN:AA52037.1; PID:9180809

A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia

R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.

Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985

A;Title: Identification and characterization of the human type II collagen gene (COL2A1).

A;Reference number: A02858; MUID:85190534; PMID:3857598

A;Accession: A02858

A;Molecule type: DNA

A;Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>

A;Cross-references: GB:J001116; NID:9180395; PIDN:AAA51997.1; PID:9180396

R;Elioma, K.; Vuorio, T.; Vuorio, E.

Nucleic Acids Res. 15, 9499-9504, 1987

A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) collagen gene

A;Reference number: A27280; MUID:88067771; PMID:2825137

A;Accession: A27280

A;Molecule type: DNA; mRNA

A;Residues: 1175-1487 <ELI>

A;Cross-references: EMBL:X06268; NID:930096; PIDN:CAA25604.1; PID:930097

A;Experimental source: fetal epiphyseal cartilage

R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.

Biochem. J. 237, 923-925, 1986

A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.

A:Reference number: A57033; MUID:87099927; PMID:3800925
 A:Accession: A57033
 A:Molecule type: protein
 A:Residues: 'XS', 1244-1246, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408 <VAN>
 A>Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
 R; Strom, C.M.; Upholt, W.B.
 Nucleic Acids Res. 12, 1025-1038, 1984
 A>Title: Isolation and characterization of genomic clones corresponding to the human type
 A:Reference number: A21733; MUID:84118798; PMID:6320112
 A:Accession: A21733
 A:Molecule type: DNA
 A:Residues: 1245-1295 <STR1>
 A:Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
 A:Accession: B21733
 A:Molecule type: DNA
 A:Residues: 894-909, 'PE' <STR2>
 A:Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
 R; Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
 Biochemistry 24, 6343-6348, 1985
 A>Title: Isolation and partial characterization of genomic clones coding for a human pro
 Gene.
 A:Reference number: A24561; MUID:86104139; PMID:3002437
 A:Accession: A24561
 A:Molecule type: DNA
 A:Residues: 1296-1358 <NUN2>
 A:Cross-references: GB:M12048; NID:g180017
 A>Note: This translation is not annotated in GenBank entry HUMCCT2A, release 111.0
 A:Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
 R; Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
 Nucleic Acids Res. 13, 2207-2225, 1985
 A>Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
 A:Reference number: I37249; MUID:85215609; PMID:2987845
 A:Accession: S59491
 A:Molecule type: DNA
 A:Residues: 7-28; 'R', 99-114; 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1
 A:Accession: I84453
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 7-28 <S&N2>
 A:Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
 A:Note: the GenBank PID is based on an incorrect reading frame
 A:Accession: I37250
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 541-560 <S&N3>
 A:Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
 A:Accession: I37251

Query Match 53.4%; Score 2476; DB 1; Length 1487;
 Best Local Similarity 50.9%; Pred. No. 2.7e-116;
 Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;
 QY 1 GPPGPGTGLGPPGGRGSGRFGPCADGVAGPKGPGAGRGSPGPPAGPKGSPGCEAGRP 60
 Db 225 GNPGEPPGVGSMGPRGPPGPKGDDGAGKPGKAGRGPPGPGAGGFFGTPGLP 284
 QY 61 -----GEAGLPKAGLITGSP---GSPGP-----DGKTGPPGPA--- 90
 Db 285 GVKHGRGYPGLDGAKGEAGAGVKGSGSPGSGSPGMPGRGLPGERGRTGPAGAGAR 344
 QY 91 GQDRPGPPGPPGAGQAGVGVGFP-----GPKGAAGFPKAGRGVGPAGVCPA 141
 Db 345 GNDQGPAGPPGPPGVPAGGPGFPAGPAKAGAGTARGPEGAQGPGRGEPCTGSPGPA 404
 QY 142 -----GKDGEAGQPPGAPGAGRGCEQCPAGSPGQGLPGPAGPPGAGKCEQV 195
 Db 405 GASGNPPTGIPGAKSGAGAPGIAGAPGPPGPPDPQAGTGPGLPGQTKGPIAGFK 464
 QY 196 GD---LGAPGSPGAGEPGTGLP-----GPPGGRGSGRGGFGGADGVA 237
 Db 465 GEQGPKEGPPGAGPQAGPAGGEGKRGARGEPGPGVGIPIPPGERGAGGNGRFGQDGLA 524
 QY 238 GPKGPAGRGSGPAGPKGSPGEGAGRPGAGLPKAGLITGSPGSPGPPDKTGPFGPAGQD 297

Db 525 GPKGAPGERPSGLAGPKGANGDPGRCEPGLPGARGLTGRPGDAGPQGVGSGARGED 584
 QY 298 GRPDPGPPGAGQAGVGMFGPKGAAGEPGKAGRGVPPGPPGAVGAGKXGAGAGQPP 357
 Db 585 GRPDPGPPGAGQAGVGMFGPKGAAGEPGKAGRGVPPGPPGAVGAGKXGAGAGQPP 644
 QY 358 GPAGPAGERGEGQAGSPGQGLPGPAGPCEAGKPGCEQGVGCDLGPAGPSGPAGE--- 413
 Db 645 GPAGPAGERGEGQAGSPGQGLPGPAGPCEAGKPGCEQGVGCDLGPAGPSGPAGE--- 704
 QY 414 -----PGPTGLP-----GPPGERGPGSGSRGPPFGADGVAGPKGPA 447
 Db 705 GERGSPGAQGLGPRGLPGTGTGPKGASGPAGPPGAQGPGLQGWGGERGAAGIAGPK 764
 QY 448 GERGS-----PGPAGPKGSPGEGAGRPBAGLPGAKGLTGSP 483
 Db 765 GDRDVGKEKPEGAPGKDGKGLTGTPTGPPGAGANGKEGVEGPPGPPGAGSAGARGPGER 824
 QY 484 GSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVGMFGPKGGAAGEP-----GKA 534
 Db 825 GETGPPGTSGIAGPPGADGQFGAKGEGEAGQKGDAGAPGQGPSPGAPGPGTGTGPK 884
 QY 535 GERGVPPGPPGA-----VGPAGKDGEGAGAOQPPGPA---GPAGERGEOGPAGSPGQF 582
 Db 885 GARGAQPPGATGFPGAAGRVGPPGSGNGNPGPPGPPGSKDGPKGARGDSGPPGRAGEP 944
 QY 583 GLPGPAGPPGEGAKGPKGQVGDLAGPSPGPPAGEPGTGLPGPPGRRGGPSRGGFPAGD 642
 Db 945 GLQGPAGPPGKEGFEFGDGPDSGABGPPGQCLAGQGVGLPGQGRGERGPPGLPGSPGE 1004
 QY 643 GVAGPKGPAGERSGPPGAPKSGSPGEGAGRPGEAGLPGA-----KGLT 684
 Db 1005 GQQGAPGASGDRGPPGVGPPGLTGPAGEFREGSPGADGPPGDRGAGVKGDRGETGAV 1064
 QY 685 GSPGSPDCKTGPSPG-----AGQDRGPPGPPGARGQAGVGMFGPKGAAGEP 735
 Db 1065 GAPGAPGPPSGPAGTGTQDRGEAGAOQPMGSPGAPAGTGGQGPGRGDKGEAGEP 1124
 QY 736 GK---AGERGVPGPAGVAGKDGEGAGAOQPPGPPGAPGAGRGBOGPAGSPGFOGLPAPA 792
 Db 1125 GERGLKGHRGFTGLQGLPGPPGSDQASGSPAGSPGPRGPPGVPVPSGDKGANGIPGPI 1184
 QY 793 GPPGEGAKPGEQGVGDLGAPGSPGAPG 820
 Db 1185 GPPGPRGRSGTGTGAPGPPGPPGPPG 1212

RESULT 6

A41182
 collagen alpha 1(II) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
 C:Accession: A41182; A44885
 J. Biol. Chem. 266, 16862-16869, 1991
 A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 A:Reference number: A41182; MUID:91359489; PMID:1885613
 A:Accession: A41182
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-1419 <MET>
 A:Cross-references: GB:M65161
 R; Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
 Development 111, 945-953, 1991
 A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage
 A:Reference number: A44885; MUID:91347939; PMID:1879363
 A:Accession: A44885
 A:Molecule type: DNA
 A:Residues: 1-28 <CH>
 A:Cross-references: GB:S63190; NID:g234368; PIDN:AA819627.1; PID:g234369
 A:Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:63192)
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Db 885 GARGAQQPPGATGTFGAAGRVGPPGANGNPGPAGPPGAGKDGKPKGVGRDGGPPGCRAGDP 944
Qy 583 GLPDPAGPGBGAKGCEQGVGPDILGAPGSPGAPGPGTGLPGPPGEGGSGRSGFPDAD 642
Db 945 GLQSPAGAPGKSGEPGDDGSLDGGPPGQGLAGQGVGLPGQGRGRTFGLPGSPGEP 1004
Qy 643 GVAGPKGPAGRGSGPAGPKGSGPCEAGRPCEAGLPGA-----KGLT 684
Db 1005 GKQAPGASGDRGPPGVPGLTGPAGBPGREGSFGADGPPRGGAAGVKGDRGTGAL 1064
Qy 685 GSPGSGPDGKTGPPGP-----AGQDRGPPGPPGARGQAGVGMFFGPKGAAGEP 735
Db 1065 GAPGAPGPPGSGPAGPTGKQDRCEAGAQAQPMGSPGAGARGIAGPQPRGDKGESGEQ 1124
Qy 736 GK---AGRGVPPGAVGAPGAGKDEAGAAQAGPPGAPGAGRGEGGAGSGSPGQGLPCPA 792
Db 1125 GERGLKGRGTGLQGLPGPPGSDQAGSAGPSPGPPGPPVGPSPGDKGNSGIPGPI 1184
Qy 793 GPPGAGKPGEGQVPGDILGAPGSPGAG 820
Db 1185 GPPGPRGRSGTGTGPPGSGPPGPPG 1212

RESULT 8
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B40333
R:Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: B40333
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1486 <SUA>
A:CROSS-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595
C:Superfamily: collagen alpha 1(i) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F:1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 51.7%; Score 2401; DB 1; Length 1486;
Best Local Similarity 48.9%; Pred. No. 1.4e-112;
Matches 493; Conservative 48; Mismatches 276; Indels 192; Gaps 16;

Qy 1 GPPGPGPTGLPGPPGEGGSGRSGFPAGDGVAGPKGP-----AGER 42
Db 209 GPRGPPGTGAPPGPGFOGNPGEPGEPGAGGPMGPRGPPGSGKPGDDGEGAKPKSGER 268
Qy 43 GSPGAPGPKGSPGEAGRP-----GEAGLPKAK---GLTGSPGSPGP--- 80
Db 269 GPPGPGAGRGPTGLPGVKGHRGYPGLDGAKGGAAGAKGEGATGAGSGPQMPGR 328
Qy 81 -----DKTGPAGQDGRPPGPPGARGQAGVGMFFGPKP 117
Db 329 GLPGRGRPGSSGAAGARGNDGLPGPAGPPGVPAGAPGPPGADGSKGEAGTGAARGE 388
Qy 118 GAAGEPKAGRGVPPGCAVCPAGKDEAGAAQAGPPGAPGAGRGEGGAGSGSPGQGLP 177
Db 389 GAQGRGSGTGPSPGAPAGSNPTDGIPLGAKSGSGFGIAGAPFPFGPPGPPQAGT 448
Qy 178 GPAGPPGAGKPE-----GEQGVPGDILGAPGSPS---GPA-----GEPGFTGLPDP 219
Db 449 GPLGPKGTGDPVAGFKGEQPKGEISAGPQAGPAGBEGKKGAGFGAAGNPP 508
Qy 220 GERGPGSRGPPGADGVAGPKGAPGERSGPPGAPKSGPGEAGRPGEAGLPAKGLTGP 279
Db 509 GERGAPGNRGPPQDGLAGPKGAPGERGVPLGCGKKGNDGPRGEPGLPCARGLTGRP 568
Qy 280 GSPGPDGKTGPPGAPQDGRPPGPPGARGQAGVGMFFGPKGAAGBEGKAGRGVPPPP 339

Db 569 GDAGPQGVKPSGASGEDRPGPPGPOGARGQPCVMGPPGPGKANGRPGKAGKGLVGP 628
Qy 340 GAVGAPAKDGEAGAQGPDPGAPGAGERGEGQGPAGSPGQGLPGPAGPPGEGAKGEGQVVP 399
Db 629 GLRLPKDKGTGSGQPNGPAGPAGERGEGQGPSPGQGLPGPPGSGEGKPGDQVPP 688
Qy 400 GDLCAQPSGPPAGS---PGPTGLPGPGERGSGSRGPPGADGVAGPKGPAGERSGCPA 456
Db 689 GEAGAPLIVGRBGRGPPGERSGSPQGLQPRGLGTPGTDTGPKGASGSPGPKGAQGP 748
Qy 457 GPKGSPGEAGRPGBAGLPGAKGLTSGSPGPDGKTGPPGAPGQDRGPPGPPGPPGCAQQA 516
Db 749 GLQMPGERGAAGISGPKGDRDGTGKXGPEGASCKDSRGLTGTGTPGPPGAPNGXKES 808
Qy 517 GVMGPPGPKGAAGBPGKAGRGVPGPGAVGPAQK-----GEAGAQGP 561
Db 809 GPSGPPGIVGARGAPGDRGNGPPGPAFAGPPGADGQSLKQDGSQKGDAGAPGP 868
Qy 562 GPA-----GPAGERGEGGAGSGPQGL-----PGPAGPGEAGKP 597
Db 869 GPSGAPGPGTGVFGPKGARGAQGPAGATGFPGAAGRVGTPGNGNPPGPPGPPGSAK 928
Qy 598 GEOGVPGDILGAP-----GPSGAPGPPGPTGL----- 623
Db 929 GPKGVGDAGPPGRAGDPGLQGAAGAPGKGEPEGEDGSPGDPGPPQGLSGQRIVGLP 988
Qy 624 -----PPGPPGRRGPPGRRGFPAGDGVAGPKGPAGERSGPPGAPKSGPGEAGRPGE- 674
Db 989 QORGERFPGLPGPSGPKGQKGGSSGDRGPPGPV---GPPGLTGSSEGEPPGPNPGSD 1045
Qy 675 -----AGLPKAGLGTGSPSGPDGKTGPPGAPGQDRGPPGPP 714
Db 1046 GPPRGDAGTIKDRGTGPIGAPGAPGAPGASVGTGKQDRGSGGPPGLGSPGA 1105
Qy 715 GARGQAGVMPGPPGKGAAGEPKAGE---RGVPPGPAVGPAGKDGAGAGQPPGPPGPA 771
Db 1106 GARGLAGPQPRGDKGEGAGERGQKGRFTGLQGLPGPPGAGDQATGAPGAPGR 1165
Qy 772 GERGEGQAGSPGQGLPGPAGPGEAGKPGEQGQVPGDLGAPGSPGAG 820
Db 1166 GPPGVPVSGKDGSGNGISGPIGPPGPRGSGTGPSPGPPGQPPGPPG 1214

RESULT 9
A40333
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: A40333
R:Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: A40333
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1492 <SUA>
A:CROSS-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63596
A:Note: This sequence is presented as substitutions relative to another sequence in a file
as they replace; the appropriate interpretation of the sequence figure was reconstructed
C:Superfamily: collagen alpha 1(i) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 51.5%; Score 2389; DB 2; Length 1492;
Best Local Similarity 50.6%; Pred. No. 5.7e-112;
Matches 477; Conservative 59; Mismatches 275; Indels 132; Gaps 14;

Qy 1 GPPGEPGPTGLPGPPGRRGPPGSGRFPDAD-----GVAGPKGAPGERSGPPGAPK 51
Db 272 GPPGQAGRGFPPTGLPGVKGHRGYPGLDGGKGAAGAKGSGASGAGAGPMPGR 331

Db 138 GAPGPMGPRGAPGERGRP---GLPGAAGARGNDGARGSDGQPPGPPGTAGTFFGSPGCAK 194
Qy 103 GARGQAGVWGPPGPKGAAGPECKAGRGVPPPP-----GAVGPA---GKGEA 147
Db 195 GEVGPAGSPGSSGAPGQGERGEPQCHAGAPGPPGPGSDGSPGKGGEMGAPGAPGLI 254
Qy 148 GAOGPPGAPGAGEGEGPAGSPGFQGLPGAPGPPGAGKPGEGVPGDGLGAPGSPGA 207
Db 255 GARGPPGPPGTVGVPQGAAGEPKKAGKADGPPRGERGAGSPGIAGPKGEDGKQSP 314
Qy 208 GEPGPTGLPFPGERGSGRGFPAGDGVAGPKGAPAGERGSPGAPKXGSPGEAGRPGEA 267
Db 315 GEPGANGLPAAAGRGVFRGPGACANGLPCKEKGPPGDRGGPPGAPRGVAGERGNL 374
Qy 268 GLPGAAGLTGSPGSPGPPGKTKTGPAGQDQRPDPGPPGARGQAGVGMFFGPKGAAGEP 327
Db 375 GGPGLRGIPGSPGSGNKPKPPGSGQGETGRGPPGSPGPRGQPPGVMGFFGPKGNDGAP 434
Qy 328 GKAGRGVPPGPAVGPAGKDEAGAGCPGAPGAPAGERGQGPAGSPGFLGPGAPGP 387
Db 435 GKNGERGPGGPGPPGQGPAGKNGETGPQGPPTGSPGDKGDTGPPGQGLGTPGTSPP 494
Qy 388 GEAGKPGQGVPGDLCAP-----GPSGAPGEPGPTGLPGERGSGGSGRGFPAGDGA 441
Db 495 GENGKPGEPGPKGEAGAGPICKGKDSGAPGERGPPGAGGPPGKGGAGPPGPEGKGA 554
Qy 442 GPKGAPGRGSP---GAPGPKGSPGEAGRPGEAGLPGAKLTGSPGSPGPKGT---GPP 495
Db 555 GPPGPPGSAAGTFLQGMFGERGPGGPKGDKGEPGSSGVDGAPKDGPRGPTPIGPP 614
Qy 496 GPAGQDGR-----PQPPGPPGARGQ---AGVWGPPGPKGAGEPKGAGERGVP 540
Db 615 GPAGQPGDKGSGAPGVFCIAGPRGGPGERGEGQPPGAPGQNGEPGCAKGERGAP 674
Qy 541 -----GPPGAVGPAGKDEAGAGQPP-----GPAAPAGERGEQ 573
Db 675 GEKGGGPPGAGPAGGSGPAGPPGQGVKGERGSPGPGGAAGFPGGRGPPGPPGSGNPN 734
Qy 574 GPAGSPFGQLGPPGAPGPEAGKPEQGV-----PGDLGAPGSPGAPGEPGPTGL- 623
Db 735 GPPGSSGAGKDGPPGPPGSGNAGAPSGPISGPKGSDGPPGERGAPGPPGAPGLGIA 794
Qy 624 -----PQPPGERGPPGSRFPAGDGVAGPKGAPAGERGSDP-----AGPKSP 666
Db 795 GLTGARGIAGPPGMPGARGSPGQIKGNGKPPGSPQNGERGPPGQGLPLGATAGEP 854
Qy 667 GBAGRPGEAGLPGA-----KGLTSGSPGPDGKTGPPPP-----AGQDGRGPP 711
Db 855 GRDGNPGSDGLPDRDGAAGKADRGKSGSPGAPGAPGHPGPPGVPAGKSGDRGETGA 914
Qy 712 GPPGARGQAGVGMFFPKGAAGEPKGAGERGV-----PGPPGAVGPAGKDEAGAG 762
Db 915 GPSGAPGAGSRGPPGPPGQPRGDKGETGERGAMGKKGHRGFGNFGAPGSPGACHQAV 974
Qy 763 GPPGAPGAGERGEGGPPGFPQGLPGAPGPPGAGKPE---GCVPGDLGAPGSPGA 819
Db 975 GSPGAPGPRGVPGSPGPKDGAASGHPGPIGPPRGNRGERGSGSPGHPGQPPGPP 1034
Qy 820 G 820
Db 1035 G 1035

RESULT 11

S59856

collagen alpha 1(III) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 09-Jul-2004

C:Accession: S59856; S62120; S16373

R:Toman, P.D.: de Crombrughe, B.

Gene 147, 161-168, 1994

A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA

A:Reference number: S59856; MUID:95011609; PMID:7926795

A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:Cross-references: UNIPROT:P08121; EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, 'G', 868-1464 <TOM>
A:Cross-references: EMBL:X52046; MUID:9575321; PIDN:CAA36279.1; PID:9575322
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16373; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57983; MUID:950476; PIDN:CAA41048.1; PID:950477

C:Genetics:
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 291/3; 309/3; 327/3; 345/3; 363/3; 381/3; 399/3; 417/3; 435/3; 453/3; 471/3; 489/3; 507/3; 525/3; 543/3; 561/3; 579/3; 597/3; 615/3; 633/3; 651/3; 669/3; 687/3; 705/3; 723/3; 741/3; 759/3; 777/3; 795/3; 813/3; 831/3; 849/3; 867/3; 885/3; 903/3; 921/3; 939/3; 957/3; 975/3; 993/3; 1011/3; 1029/3; 1047/3; 1065/3; 1083/3; 1101/3; 1119/3; 1137/3; 1155/3; 1173/3; 1191/3; 1209/3; 1227/3; 1245/3; 1263/3; 1281/3; 1299/3; 1317/3; 1335/3; 1353/3; 1371/3; 1389/3; 1407/3; 1425/3; 1443/3; 1461/3; 1479/3; 1497/3; 1515/3; 1533/3; 1551/3; 1569/3; 1587/3; 1605/3; 1623/3; 1641/3; 1659/3; 1677/3; 1695/3; 1713/3; 1731/3; 1749/3; 1767/3; 1785/3; 1803/3; 1821/3; 1839/3; 1857/3; 1875/3; 1893/3; 1911/3; 1929/3; 1947/3; 1965/3; 1983/3; 2001/3; 2019/3; 2037/3; 2055/3; 2073/3; 2091/3; 2109/3; 2127/3; 2145/3; 2163/3; 2181/3; 2199/3; 2217/3; 2235/3; 2253/3; 2271/3; 2289/3; 2307/3; 2325/3; 2343/3; 2361/3; 2379/3; 2397/3; 2415/3; 2433/3; 2451/3; 2469/3; 2487/3; 2505/3; 2523/3; 2541/3; 2559/3; 2577/3; 2595/3; 2613/3; 2631/3; 2649/3; 2667/3; 2685/3; 2703/3; 2721/3; 2739/3; 2757/3; 2775/3; 2793/3; 2811/3; 2829/3; 2847/3; 2865/3; 2883/3; 2901/3; 2919/3; 2937/3; 2955/3; 2973/3; 2991/3; 3009/3; 3027/3; 3045/3; 3063/3; 3081/3; 3099/3; 3117/3; 3135/3; 3153/3; 3171/3; 3189/3; 3207/3; 3225/3; 3243/3; 3261/3; 3279/3; 3297/3; 3315/3; 3333/3; 3351/3; 3369/3; 3387/3; 3405/3; 3423/3; 3441/3; 3459/3; 3477/3; 3495/3; 3513/3; 3531/3; 3549/3; 3567/3; 3585/3; 3603/3; 3621/3; 3639/3; 3657/3; 3675/3; 3693/3; 3711/3; 3729/3; 3747/3; 3765/3; 3783/3; 3801/3; 3819/3; 3837/3; 3855/3; 3873/3; 3891/3; 3909/3; 3927/3; 3945/3; 3963/3; 3981/3; 3999/3; 4017/3; 4035/3; 4053/3; 4071/3; 4089/3; 4107/3; 4125/3; 4143/3; 4161/3; 4179/3; 4197/3; 4215/3; 4233/3; 4251/3; 4269/3; 4287/3; 4305/3; 4323/3; 4341/3; 4359/3; 4377/3; 4395/3; 4413/3; 4431/3; 4449/3; 4467/3; 4485/3; 4503/3; 4521/3; 4539/3; 4557/3; 4575/3; 4593/3; 4611/3; 4629/3; 4647/3; 4665/3; 4683/3; 4701/3; 4719/3; 4737/3; 4755/3; 4773/3; 4791/3; 4809/3; 4827/3; 4845/3; 4863/3; 4881/3; 4899/3; 4917/3; 4935/3; 4953/3; 4971/3; 4989/3; 5007/3; 5025/3; 5043/3; 5061/3; 5079/3; 5097/3; 5115/3; 5133/3; 5151/3; 5169/3; 5187/3; 5205/3; 5223/3; 5241/3; 5259/3; 5277/3; 5295/3; 5313/3; 5331/3; 5349/3; 5367/3; 5385/3; 5403/3; 5421/3; 5439/3; 5457/3; 5475/3; 5493/3; 5511/3; 5529/3; 5547/3; 5565/3; 5583/3; 5601/3; 5619/3; 5637/3; 5655/3; 5673/3; 5691/3; 5709/3; 5727/3; 5745/3; 5763/3; 5781/3; 5799/3; 5817/3; 5835/3; 5853/3; 5871/3; 5889/3; 5907/3; 5925/3; 5943/3; 5961/3; 5979/3; 5997/3; 6015/3; 6033/3; 6051/3; 6069/3; 6087/3; 6105/3; 6123/3; 6141/3; 6159/3; 6177/3; 6195/3; 6213/3; 6231/3; 6249/3; 6267/3; 6285/3; 6303/3; 6321/3; 6339/3; 6357/3; 6375/3; 6393/3; 6411/3; 6429/3; 6447/3; 6465/3; 6483/3; 6501/3; 6519/3; 6537/3; 6555/3; 6573/3; 6591/3; 6609/3; 6627/3; 6645/3; 6663/3; 6681/3; 6699/3; 6717/3; 6735/3; 6753/3; 6771/3; 6789/3; 6807/3; 6825/3; 6843/3; 6861/3; 6879/3; 6897/3; 6915/3; 6933/3; 6951/3; 6969/3; 6987/3; 7005/3; 7023/3; 7041/3; 7059/3; 7077/3; 7095/3; 7113/3; 7131/3; 7149/3; 7167/3; 7185/3; 7203/3; 7221/3; 7239/3; 7257/3; 7275/3; 7293/3; 7311/3; 7329/3; 7347/3; 7365/3; 7383/3; 7401/3; 7419/3; 7437/3; 7455/3; 7473/3; 7491/3; 7509/3; 7527/3; 7545/3; 7563/3; 7581/3; 7599/3; 7617/3; 7635/3; 7653/3; 7671/3; 7689/3; 7707/3; 7725/3; 7743/3; 7761/3; 7779/3; 7797/3; 7815/3; 7833/3; 7851/3; 7869/3; 7887/3; 7905/3; 7923/3; 7941/3; 7959/3; 7977/3; 7995/3; 8013/3; 8031/3; 8049/3; 8067/3; 8085/3; 8103/3; 8121/3; 8139/3; 8157/3; 8175/3; 8193/3; 8211/3; 8229/3; 8247/3; 8265/3; 8283/3; 8301/3; 8319/3; 8337/3; 8355/3; 8373/3; 8391/3; 8409/3; 8427/3; 8445/3; 8463/3; 8481/3; 8499/3; 8517/3; 8535/3; 8553/3; 8571/3; 8589/3; 8607/3; 8625/3; 8643/3; 8661/3; 8679/3; 8697/3; 8715/3; 8733/3; 8751/3; 8769/3; 8787/3; 8805/3; 8823/3; 8841/3; 8859/3; 8877/3; 8895/3; 8913/3; 8931/3; 8949/3; 8967/3; 8985/3; 9003/3; 9021/3; 9039/3; 9057/3; 9075/3; 9093/3; 9111/3; 9129/3; 9147/3; 9165/3; 9183/3; 9201/3; 9219/3; 9237/3; 9255/3; 9273/3; 9291/3; 9309/3; 9327/3; 9345/3; 9363/3; 9381/3; 9399/3; 9417/3; 9435/3; 9453/3; 9471/3; 9489/3; 9507/3; 9525/3; 9543/3; 9561/3; 9579/3; 9597/3; 9615/3; 9633/3; 9651/3; 9669/3; 9687/3; 9705/3; 9723/3; 9741/3; 9759/3; 9777/3; 9795/3; 9813/3; 9831/3; 9849/3; 9867/3; 9885/3; 9903/3; 9921/3; 9939/3; 9957/3; 9975/3; 9993/3; 10011/3; 10029/3; 10047/3; 10065/3; 10083/3; 10101/3; 10119/3; 10137/3; 10155/3; 10173/3; 10191/3; 10209/3; 10227/3; 10245/3; 10263/3; 10281/3; 10299/3; 10317/3; 10335/3; 10353/3; 10371/3; 10389/3; 10407/3; 10425/3; 10443/3; 10461/3; 10479/3; 10497/3; 10515/3; 10533/3; 10551/3; 10569/3; 10587/3; 10605/3; 10623/3; 10641/3; 10659/3; 10677/3; 10695/3; 10713/3; 10731/3; 10749/3; 10767/3; 10785/3; 10803/3; 10821/3; 10839/3; 10857/3; 10875/3; 10893/3; 10911/3; 10929/3; 10947/3; 10965/3; 10983/3; 11001/3; 11019/3; 11037/3; 11055/3; 11073/3; 11091/3; 11109/3; 11127/3; 11145/3; 11163/3; 11181/3; 11199/3; 11217/3; 11235/3; 11253/3; 11271/3; 11289/3; 11307/3; 11325/3; 11343/3; 11361/3; 11379/3; 11397/3; 11415/3; 11433/3; 11451/3; 11469/3; 11487/3; 11505/3; 11523/3; 11541/3; 11559/3; 11577/3; 11595/3; 11613/3; 11631/3; 11649/3; 11667/3; 11685/3; 11703/3; 11721/3; 11739/3; 11757/3; 11775/3; 11793/3; 11811/3; 11829/3; 11847/3; 11865/3; 11883/3; 11901/3; 11919/3; 11937/3; 11955/3; 11973/3; 11991/3; 12009/3; 12027/3; 12045/3; 12063/3; 12081/3; 12099/3; 12117/3; 12135/3; 12153/3; 12171/3; 12189/3; 12207/3; 12225/3; 12243/3; 12261/3; 12279/3; 12297/3; 12315/3; 12333/3; 12351/3; 12369/3; 12387/3; 12405/3; 12423/3; 12441/3; 12459/3; 12477/3; 12495/3; 12513/3; 12531/3; 12549/3; 12567/3; 12585/3; 12603/3; 12621/3; 12639/3; 12657/3; 12675/3; 12693/3; 12711/3; 12729/3; 12747/3; 12765/3; 12783/3; 12801/3; 12819/3; 12837/3; 12855/3; 12873/3; 12891/3; 12909/3; 12927/3; 12945/3; 12963/3; 12981/3; 13000/3; 13018/3; 13036/3; 13054/3; 13072/3; 13090/3; 13108/3; 13126/3; 13144/3; 13162/3; 13180/3; 13198/3; 13216/3; 13234/3; 13252/3; 13270/3; 13288/3; 13306/3; 13324/3; 13342/3; 13360/3; 13378/3; 13396/3; 13414/3; 13432/3; 13450/3; 13468/3; 13486/3; 13504/3; 13522/3; 13540/3; 13558/3; 13576/3; 13594/3; 13612/3; 13630/3; 13648/3; 13666/3; 13684/3; 13702/3; 13720/3; 13738/3; 13756/3; 13774/3; 13792/3; 13810/3; 13828/3; 13846/3; 13864/3; 13882/3; 13900/3; 13918/3; 13936/3; 13954/3; 13972/3; 13990/3; 14008/3; 14026/3; 14044/3; 14062/3; 14080/3; 14098/3; 14116/3; 14134/3; 14152/3; 14170/3; 14188/3; 14206/3; 14224/3; 14242/3; 14260/3; 14278/3; 14296/3; 14314/3; 14332/3; 14350/3; 14368/3; 14386/3; 14404/3; 14422/3; 14440/3; 14458/3; 14476/3; 14494/3; 14512/3; 14530/3; 14548/3; 14566/3; 14584/3; 14602/3; 14620/3; 14638/3; 14656/3; 14674/3; 14692/3; 14710/3; 14728/3; 14746/3; 14764/3; 14782/3; 14800/3; 14818/3; 14836/3; 14854/3; 14872/3; 14890/3; 14908/3; 14926/3; 14944/3; 14962/3; 14980/3; 14998/3; 15016/3; 15034/3; 15052/3; 15070/3; 15088/3; 15106/3; 15124/3; 15142/3; 15160/3; 15178/3; 15196/3; 15214/3; 15232/3; 15250/3; 15268/3; 15286/3; 15304/3; 15322/3; 15340/3; 15358/3; 15376/3; 15394/3; 15412/3; 15430/3; 15448/3; 15466/3; 15484/3; 15502/3; 15520/3; 15538/3; 15556/3; 15574/3; 15592/3; 15610/3; 15628/3; 15646/3; 15664/3; 15682/3; 15700/3; 15718/3; 15736/3; 15754/3; 15772/3; 15790/3; 15808/3; 15826/3; 15844/3; 15862/3; 15880/3; 15898/3; 15916/3; 15934/3; 15952/3; 15970/3; 15988/3; 16006/3; 16024/3; 16042/3; 16060/3; 16078/3; 16096/3; 16114/3; 16132/3; 16150/3; 16168/3; 16186/3; 16204/3; 16222/3; 16240/3; 16258/3; 16276/3; 16294/3; 16312/3; 16330/3; 16348/3; 16366/3; 16384/3; 16402/3; 16420/3; 16438/3; 16456/3; 16474/3; 16492/3; 16510/3; 16528/3; 16546/3; 16564/3; 16582/3; 16600/3; 16618/3; 16636/3; 16654/3; 16672/3; 16690/3; 16708/3; 16726/3; 16744/3; 16762/3; 16780/3; 16798/3; 16816/3; 16834/3; 16852/3; 16870/3; 16888/3; 16906/3; 16924/3; 16942/3; 16960/3; 16978/3; 16996/3; 17014/3; 17032/3; 17050/3; 17068/3; 17086/3; 17104/3; 17122/3; 17140/3; 17158/3; 17176/3; 17194/3; 17212/3; 17230/3; 17248/3; 17266/3; 17284/3; 17302/3; 17320/3; 17338/3; 17356/3; 17374/3; 17392/3; 17410/3; 17428/3; 17446/3; 17464/3; 17482/3; 17500/3; 17518/3; 17536/3; 17554/3; 17572/3; 17590/3; 17608/3; 17626/3; 17644/3; 17662/3; 17680/3; 17698/3; 17716/3; 17734/3; 17752/3; 17770/3; 17788/3; 17806/3; 17824/3; 17842/3; 17860/3; 17878/3; 17896/3; 17914/3; 17932/3; 17950/3; 17968/3; 17986/3; 18004/3; 18022/3; 18040/3; 18058/3; 18076/3; 18094/3; 18112/3; 18130/3; 18148/3; 18166/3; 18184/3; 18202/3; 18220/3; 18238/3; 18256/3; 18274/3; 18292/3; 18310/3; 18328/3; 18346/3; 18364/3; 18382/3; 18400/3; 18418/3; 18436/3; 18454/3; 18472/3; 18490/3; 18508/3; 18526/3; 18544/3; 18562/3; 18580/3; 18598/3; 18616/3; 18634/3; 18652/3; 18670/3; 18688/3; 18706/3; 18724/3; 18742/3; 18760/3; 18778/3; 18796/3; 18814/3; 18832/3; 18850/3; 18868/3; 18886/3; 18904/3; 18922/3; 18940/3; 18958/3; 18976/3; 18994/3; 19012/3; 19030/3; 19048/3; 19066/3; 19084/3; 19102/3; 19120/3; 19138/3; 19156/3; 19174/3; 19192/3; 19210/3; 19228/3; 19246/3; 19264/3; 19282/3; 19300/3; 19318/3; 19336/3; 19354/3; 19372/3; 19390/3; 19408/3; 19426/3; 19444/3; 19462/3; 19480/3; 19498/3; 19516/3; 19534/3; 19552/3; 19570/3; 19588/3; 19606/3; 19624/3; 19642/3; 19660/3; 19678/3; 19696/3; 19714/3; 19732/3; 19750/3; 19768/3; 19786/3; 19804/3; 19822/3; 19840/3; 19858/3; 19876/3; 19894/3; 19912/3; 19930/3; 19948/3; 19966/3; 19984/3; 20002/3; 20020/3; 20038/3; 20056/3; 20074/3; 20092/3; 20110/3; 20128/3; 20146/3; 20164/3; 20182/3; 20200/3; 20218/3; 20236/3; 20254/3; 20272/3; 20290/3; 20308/3; 20326/3; 20344/3; 20362/3; 20380/3; 20398/3; 20416/3; 20434/3; 20452/3; 20470/3; 20488/3; 20506/3; 20524/3; 20542/3; 20560/3; 20578/3; 20596/3; 20614/3; 20632/3; 20650/3; 20668/3; 20686/3; 20704/3; 20722/3; 20740/3; 20758/3; 20776/3; 20794/3; 20812/3; 20830/3; 20848/3; 20866/3; 20884/3; 20902/3; 20920/3; 20938/3; 20956/3; 20974/3; 20992/3; 21010/3; 21028/3; 21046/3; 21064/3; 21082/3; 21100/3; 21118/3; 21136/3; 21154/3; 21172/3; 21190/3; 21208/3; 21226/3; 2124

Db 758 GP---RGAPGIGPPGP---AGQPDGKGGSPGLPGIAGPRGPGRGHGGPPGPPAGFP 811

QY 565 -----GPAGRGGCGPAGSPGQGLPGPAGPPGE---AGKPGGQGVPGDGLGAPGPPGPA 615

Db 812 GAPGQNGPFGAKGGAPEKGGEGGPPGPGAGFTGSSGPAGPPGQGVKGRSGSPGPGTA 871

QY 616 GEPGPTGLPGPPGRRGGSGSFGFCADVAGPKPAGRGSPGPGAPKPGSPGGRGPRGE- 674

Db 872 GFPGRGLPGPPGNNGNPPGPPGAPGKDGKPPGAGNSGSPGNPGIAGPKGADGAPGK 931

QY 675 -----AGLPGAKGLTGSPGSPGPDGKTGP-----PGPAGQDGRP 708

Db 932 GPPCAQGPSPGPGIAGTLGARGLAGPCMPGPRSPGPGQGTGKSGRPGASGNGR 991

QY 709 GPPGP-----FGARGQAGVMGFPKPKGAAGE---PGKAGRGVFPGP 747

Db 992 GPPGPGQGLPGQGTAGTGFGRDNGSDGQPGKGDGKDRGNSGFGAPGAPGHPGP 1051

QY 748 GAVGPAGKDG------AGAQPFGPAGPAGRGEGQPGSGFGQGLPGPA 792

Db 1052 GPVPSGKSGDRGTGPAGPSGAPGAPAGAPGQPGRGDKGTGTRGSGNGIKGHRGP 1111

QY 793 GPPGEAGKPGGQGVPGDGLGAPGSPGAP 820

Db 1112 GNPFGPGSGAAGHGAIGSPGAPRG 1139

RESULT 12

CGHUTL

collagen alpha 1(III) chain precursor - human

N;Alternate names: procollagen alpha 1(III) chain

C;Species: Homo sapiens (man)

C;Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004

C;Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399

R;Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A;Reference number: S05272

A;Accession: S05272

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1240,'V',1242-1466 <PRC>

A;Cross-references: UNIPROT:P02461; EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058

R;Aila-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A;Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human procollagen.

A;Reference number: S04642; MUID:89350838; PMID:2764886

A;Accession: S04642

A;Molecule type: mRNA

A;Residues: 1-1196 <ALA>

A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058

A;Note: the complete sequence is not shown

R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene

A;Reference number: PE0011; MUID:89378752; PMID:2777083

A;Accession: PE0011

A;Molecule type: DNA

A;Residues: 1-176 <BEN>

A;Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814

R;Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human procollagen type III

A;Reference number: S01726; MUID:88303360; PMID:3405773

A;Accession: S01726

A;Molecule type: mRNA

A;Residues: 1-170 <TOM>

A;Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061

A;Note: the authors translated the codon CAG for residue 154 as His

R;Janeczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1989

A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.

A;Reference number: S04887; MUID:89386015; PMID:2780304

A;Accession: S04887

A;Molecule type: mRNA

A;Residues: 149-163,'G',164-240,'D',242-471,'D',473-487,'L',489,'S',491-613,'Y',615-634,'A';

A;Cross-references: EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:g930045

A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide

R;Seyer, J.M.; Kang, A.H.

Biochemistry 16, 1158-1164, 1977

A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides

A;Reference number: A90399; MUID:7134724; PMID:557335

A;Accession: A90399

A;Molecule type: protein

A;Residues: 'V',169-225,229-232,'P',234-292,'D',294-398 <SEY1>

A;Experimental source: liver

A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose

R;Seyer, J.M.

submitted to the Atlas, December 1977

A;Reference number: A94562

A;Accession: A94562

A;Molecule type: protein

A;Residues: 'V',169-225,229-277,'A',279-292,'D',294,'S',296-398 <SEY2>

A;Experimental source: liver

A;Note: author submitted corrections to A90399

R;Milwicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.

Am. J. Hum. Genet. 53, 62-70, 1993

A;Title: Parental somatic and germ-line mosaicism for a multigene deletion with unusual

A;Reference number: I51868; MUID:93304430; PMID:8317500

A;Accession: I51868

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 186-194 <MIL>

A;Cross-references: GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:g4261637

R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.

Biochem. J. 311, 939-943, 1995

A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1

A;Reference number: S59511; MUID:96067614; PMID:7487954

A;Accession: S59511

A;Molecule type: mRNA

A;Residues: 302-423 <CHI>

A;Cross-references: GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:g1195577

R;Seyer, J.M.; Kang, A.H.

Biochemistry 17, 3404-3411, 1978

A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBR peptides

A;Reference number: A90414; MUID:79000343; PMID:687591

A;Accession: A90414

A;Molecule type: protein

A;Residues: 399-675,'N',677-727 <SEY3>

A;Experimental source: liver

R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.

J. Biol. Chem. 266, 5256-5259, 1991

A;Title: G to T transversion at position +5 of a splice donor site causes skipping of the

A;Reference number: I55349; MUID:91161621; PMID:1672129

A;Accession: I55349

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 537-605 <LER>

A;Cross-references: GB:M59312; NID:g180815; PIDN:AAA52041.1; PID:g180816

R;Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A;Title: Covalent structure of collagen: amino acid sequence of alpha(III)-CB5 from ty

A;Reference number: A90438; MUID:80198282; PMID:6246925

A;Accession: A90438

A;Molecule type: protein

A;Residues: 728-895,'A',897-964 <SEY4>

A;Experimental source: liver

R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan,

J. Biol. Chem. 265, 17070-17077, 1990

A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and

A;Reference number: A38303; MUID:91009133; PMID:2145268

A;Accession: A38303

A;Molecule type: mRNA

A;Residues: 861-1015 <COL>

A;Cross-references: GB:J05617; GB:M55603; NID:g180878; PIDN:AAB59383.1; PID:g1

F:154-167/Region: amino-terminal nonhelical telopeptide

F:168-1196/Region: helical

F:1091-1093/Region: cell attachment (R-G-D) motif

F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide

F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>

F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

F:124/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F:153-154/Cleavage site: Pro-Gln (procollagen N-endorpeptidase) #status predicted

F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F:161,1212/Modified site: allysine (Lys) #status predicted

F:263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental

F:263/Binding site: carbohydrate (Lys) (covalent) #status experimental

F:584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

F:948-949/Cleavage site: Gly-Ile (Collagenase) #status experimental

F:1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 51.1%; Score 2370; DB 1; Length 1466;
Best Local Similarity 49.5%; Pred. No. 4.9e-111;
Matches 476; Conservative 56; Mismatches 285; Indels 144; Gaps 16;

QY	1	GPPEPPTGLPPGERGGPSRGFP-----GADVAGPKGPAGER 42
DB	234	GRPRPGERGLPPGGIKGPAGIPGPFQMKGHRGPDGRNGEKGTGAPCLKGKENGLPGEN 293
QY	43	GSPGPAGPKGSPERAGRPGEAGLPCAKGLTSCSPSGPDGKTGTPPGAGQDRP----- 96
DB	294	GAPGMPGRGAPGERRP---GLFGAAGARGNDGARGSDGQPGPPGTATGFPGSPGAK 350
QY	97	---GPPGPPCARQAGVGMVPPGPKP---GAAGEPKKAGRGVPPGPAVGPA---GKDGEA 147
DB	351	GEVGPAGSPGSGNAPQGRGPPGPOCHAGAQQPPGPGINGSPGGKGEMGAPGIPGAPGLM 410
QY	148	GAQGPFGPAGPAGERGBQGPAGSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGA 207
DB	411	GARGPPGAPACANGAPGLRGAGGEPGKNKAGEPCPRGERGEAGIPGVPCAKDEGDKGSP 470
QY	208	GERGPTGLPPGERGGPSRGPFPCADGVAGPKGPAGERGSPGPAGPKGSPERAGRPGEA 267
DB	471	GEPCANGLPCAAGERGAPGRFPAGPAGNPIGEEKGPAGERGAPGAPGPRGAAGEPRDGPV 530
QY	268	GLPGAKGLTSCSPSGPDGKTGPPGAGQDGRGPPGPPGARGQAGVGMFPCKGAAGEP 327
DB	531	GGPGMRGMPGSPGGSDGPKPPGSPGSGESRPPGPPGSPRGQPGVMGFPGPKGNDGAP 590
QY	328	GKAGERGVPPGAVGAPGAKDGEAGAACQPPGAPGAGERGEOCPAGSPGFQGLPGPAGPP 387
DB	591	GKNGERGGPGRPQGPGRNGETGPPQGPPTGPGDKDGTGPPGQGLQGLPGTGGPP 650
QY	388	GEAKPGEOQVVPDGLAP-----GPSGPAGPGPT 417
DB	651	GENGKDEGPFGDAGAPGAPGKGDAGAPGERGPPGLAGAPLRRGAGPPGPEGGKGA 710
QY	418	GLFGPGERGGPSRGPFADGVAGPKGPAGERGSP-----GPAGPKGSP 462
DB	711	GPPGPPGAAGTGLGQMPGERGGGLSGSPGPKDKGSPGPGADGVFPKDGPRGPTGPIGPP 770
QY	463	GEAGRP---GEAGLPGAKGLTSCSPSPDCKTGPAGQDGRGPPGPPGARGQAGV- 518
DB	771	GPAGQGDKGEGAPGLPGIAGPKGSPGERGETGPPGAPFGAPGQNGERPGKGERGAP 830
QY	519	--MGFPKPGAAGEPKKAGERGVGPP-----PCAVGPAGKDGEGAGAQ 558
DB	831	GEXGEGPPVAGPPGSGSPAGPPGPGQVKGERGSPGFGAAGFPFGARGLPPPGSGNGNP 890
QY	559	GPPGPAGPAGERGQGPAGSPGFQGLPGPAGPPEAGKPGEGQVPGDGLGAPGSPGAPGP 618
DB	891	GPPGSPGSKDGPPPGAGNNTGAPGSPGVSGPKGDAGQFGEKSGSPCAQGPFGAPGLGTA 950
QY	619	GPTGL-----PGPPGERGGPSRGF-----PGADGVAGPKGPAGERGSPGAPKGPSP 666
DB	951	GITTGAGLAGPAGMPGRGSPGPGVKGESKPKGANGUSGERGPPGQGLPLGATAGP 1010
QY	667	GEAGRPGEAGLPGA-----KGLTSCSPSGPDGKTGPPGP-----AGQDGRGPP 711

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:17:54 ; Search time 121.782 Seconds
(without alignments)
3452.202 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGPGTGLPGPPGRRG.....GEGVPGDLGAPSPGPG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2765	59.6	1 CAIL1 HUMAN	P02452 homo sapien
2	2761	59.5	2 O76045	O76045 homo sapien
3	2761	59.5	2 Q8N473	Q8N473 homo sapien
4	2759	59.5	1 CAIL1 CANFA	Q9X8J7 canis fami
5	2748.5	59.2	2 Q63079	Q63079 rattus norv
6	2732.5	58.9	1 CAIL1 MOUSE	P11087 mus musculu
7	2732.5	58.9	2 Q810J9	Q810J9 mus musculu
8	2681.5	57.8	2 Q6LAN8	Q6LAN8 homo sapien
9	2679	57.7	1 CAIL1 CHICK	P02457 gallus gall
10	2568	55.3	2 O93251	O93251 rana catesb
11	2561.5	55.2	2 O9YIB4	O9YIB4 cynops pyrr
12	2558.5	55.1	2 Q640B2	Q640B2 xenopus tro
13	2529	54.5	2 Q802B5	Q802B5 xenopus lae
14	2494	53.8	2 Q6PCL3	Q6PCL3 mus musculu
15	2486	53.6	2 Q14046	Q14046 homo sapien
16	2486	53.6	2 Q14047	Q14047 homo sapien
17	2481	53.5	2 Q77753	Q77753 canis fami
18	2479	53.4	2 Q9IB91	Q9IB91 xenopus lae
19	2477	53.4	2 Q28396	Q28396 equus cabal
20	2476	53.4	1 CAIL2 HUMAN	P02458 homo sapien
21	2475.5	53.4	2 Q90W37	Q90W37 gallus gall
22	2465.5	53.1	2 Q72227	Q72227 gallus gall
23	2464	53.1	2 Q62031	Q62031 mus musculu
24	2464	53.1	2 Q62033	Q62033 mus musculu
25	2464	53.1	1 CAIL2 MOUSE	P28481 mus musculu
26	2464	53.1	1 Q62032	Q62032 mus musculu
27	2455	52.9	2 Q80VY3	Q80VY3 mus musculu
28	2455	52.9	2 Q641K3	Q641K3 mus musculu
29	2454	52.9	2 Q63123	Q63123 rattus norv
30	2449	52.8	1 Q80X38	Q80X38 mus musculu
31	2430	52.4	2 Q6P4U1	Q6P4U1 brachydania

32	2428.5	52.3	1447	2	O6ULJ5	Q6ulj5 brachydania
33	2412	52.0	1418	2	O9W7E9	O9w7e9 cynops pyrr
34	2405	51.8	1492	2	Q6P4Z2	Q6p4z2 xenopus tro
35	2401	51.7	1486	2	Q91717	Q91717 xenopus lae
36	2399	51.7	1486	2	Q7ZTI6	Q7zti6 xenopus lae
37	2390	51.5	1449	2	Q6PEI9	Q6pei9 brachydania
38	2386	51.4	1491	2	O91718	Q91718 xenopus lae
39	2385	51.4	1491	2	Q7ZTM4	Q7ztm4 xenopus lae
40	2384	51.4	1449	2	O6NZI5	O6nzi5 brachydania
41	2379	51.3	1049	1	CAI3_BOVIN	P04258 bos taurus
42	2370	51.1	1466	1	CAI3_HUMAN	P08121 homo sapien
43	2368.5	51.0	1464	1	CAI3_MOUSE	Q7tt32 mus musculu
44	2368.5	51.0	1464	2	O7TT32	Q7tt32 mus musculu
45	2368.5	51.0	1464	2	Q8BKX2	Q8bky2 mus musculu

ALIGNMENTS

RESULT 1

ID	CALL_HUMAN	STANDARD;	PRT;	1464 AA.
AC	P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176; Q15201; Q16050; Q7KZ30; Q7KZ34; Q8IVI5; Q9UML6; Q9UMW7;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	25-JAN-2005 (Rel. 46, Last annotation update)			
DE	Collagen alpha 1(I) chain precursor.			
GN	Name=COL1A1;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT.SER-1434.			
RA	Dalgleish R.;			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1-589 FROM N.A.			
RX	MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;			
RA	D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,			
RA	Pretorius P.J.;			
RT	"Complete nucleotide sequence of the region encompassing the first			
RT	twenty-five exons of the human pro alpha 1(I) collagen gene			
RL	(COL1A1).";			
RN	[3]			
RP	SEQUENCE OF 1-472 FROM N.A.			
RX	MEDLINE=89025644; PubMed=3178743;			
RA	Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,			
RA	Jaenisch R., Prockup D.J.;			
RT	"Structure of a full-length cDNA clone for the prepro alpha 1(I) chain			
RT	of human type I procollagen.";			
RL	Biochem. J. 253:919-922(1988).			
RN	[4]			
RP	SEQUENCE OF 1-181 FROM N.A.			
RX	MEDLINE=84270697; PubMed=6462220;			
RA	Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,			
RA	Myers J., Williams C., Ramirez F.;			
RT	"Human pro alpha 1(I) collagen gene structure reveals evolutionary			
RT	conservation of a pattern of introns and exons.";			
RL	Nature 310:337-340(1984).			
RN	[5]			
RP	SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.			
RC	TISSUE=Skin;			
RX	MEDLINE=71038625; PubMed=5529814;			
RA	Click E.M., Bornstein P.;			
RT	"Isolation and characterization of the cyanogen bromide peptides from			
RT	the alpha 1 and alpha 2 chains of human skin collagen.";			
RL	Biochemistry 9:4699-4706(1970).			
RN	[6]			
RP	SEQUENCE OF 425-1464 FROM N.A.			
RX	MEDLINE=84080385; PubMed=6689127;			

RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RA "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [7]
 RN SEQUENCE OF 472-607 FROM N.A.
 RP PubMed=2981843;
 RX Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
 RA "Multixon deletion in an osteogenesis imperfecta variant with
 RT increased type III collagen mRNA.";
 RL J. Biol. Chem. 260:691-694(1985).
 RN [8]
 RN SEQUENCE OF 488-625 FROM N.A.
 RP PubMed=3857621;
 RX Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
 RA "Intron-mediated recombination may cause a deletion in an alpha 1 type
 RT I collagen chain in a lethal form of osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
 RN [9]
 RN SEQUENCE OF 1179-1464 FROM N.A.; VARIANTS OI-II HIS-1277; ARG-1388 AND
 RP MEDLINE=93352646; PubMed=8349697;
 RX Chessler S.D., Wallis G.A., Byers P.H.;
 RA "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
 RT chain of type I collagen result in defective chain association and
 RT produce lethal osteogenesis imperfecta.";
 RL J. Biol. Chem. 268:18218-18225(1993).
 RN [10]
 RN SEQUENCE OF 1229-1454 FROM N.A.
 RP TISSUE=Bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekela J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [11]
 RN SEQUENCE OF 1-34 FROM N.A.
 RP MEDLINE=85130970; PubMed=2857713;
 RX Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
 RA "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 RT Promoter structure, AluI repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [12]
 RN SEQUENCE OF 1-34 FROM N.A.
 RP MEDLINE=88097389; PubMed=3480516;
 RX Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RA "Regulatory elements in the first intron contribute to transcriptional
 RT control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [13]
 RN SEQUENCE OF 1-44 FROM N.A.
 RP MEDLINE=88033098; PubMed=2822714;
 RX Rousow C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [14]
 RN SEQUENCE OF 33-52.
 RP PubMed=2318855;
 RX Wirtz M.K., Keene D.R., Hori H., Glanville R.W., Steinmann B.,
 RA Rao V.H., Hollister D.W.;
 RT "In vivo and in vitro noncovalent association of excised alpha 1 (I)
 RT amino-terminal propeptides with mutant pro alpha 2(I) collagen chains
 RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
 RT VII.";
 RL J. Biol. Chem. 265:6312-6317(1990).
 RN [15]
 RN SEQUENCE OF 156-183 FROM N.A.
 RP PubMed=2767050;
 RX Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,
 RA

RA Bateman J.F.;
 RT "A base substitution in the exon of a collagen gene causes alternative
 RT splicing and generates a structurally abnormal polypeptide in a
 RT patient with Ehlers-Danlos syndrome type VII.";
 RL EMBO J. 8:1705-1710(1989).
 RN [16]
 RN SEQUENCE OF 175-187 AND 274-289.
 RP PubMed=2169412;
 RX Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
 RA Muller P.K.;
 RT "A critical crosslink region in human-bone-derived collagen type I.
 RT Specific cleavage site at residue Leu95.";
 RL Eur. J. Biochem. 192:153-159(1990).
 RN [17]
 RN SEQUENCE OF 263-268.
 RP TISSUE=Skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected vertebrate
 RT collagens. A possible role of the carbohydrate in fibril formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [18]
 RN SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND
 RP 1143-1162 FROM N.A.
 RX PubMed=2374517;
 RA Labhard M.B., Hollister D.W.;
 RT "Segmental amplification of the entire helical and telopeptide regions
 RT of the cDNA for human alpha 1 (I) collagen.";
 RL Matrix 10:124-130(1990).
 RN [19]
 RN SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
 RP MEDLINE=83064528; PubMed=6183642;
 RX Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
 RA "Cloning and characterization of five overlapping cDNAs specific for
 RT the human pro alpha 1(I) collagen chain.";
 RL Nucleic Acids Res. 10:5925-5934(1982).
 RN [20]
 RN SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
 RP PubMed=2339700;
 RX Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
 RA "Variable expression of osteogenesis imperfecta in a nuclear family is
 RT explained by somatic mosaicism for a lethal point mutation in the
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
 RL Am. J. Hum. Genet. 46:1034-1040(1990).
 RN [21]
 RN SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
 RP MEDLINE=95187161; PubMed=7881420;
 RX Porlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
 RA Brunelli P.C., Mottes M.;
 RT "Severe (type III) osteogenesis imperfecta due to glycine
 RT substitutions in the central domain of the collagen triple helix.";
 RL Hum. Mol. Genet. 3:2201-2206(1994).
 RN [22]
 RN SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.
 RP PubMed=3170557;
 RX Conn D.H., Apone S., Eyre D.R., Starman B.J., Andreassen P.,
 RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
 RA "Substitution of cysteine for glycine within the carboxyl-terminal
 RT telopeptide of the alpha 1 chain of type I collagen produces mild
 RT osteogenesis imperfecta.";
 RL J. Biol. Chem. 263:14605-14607(1988).
 RN [23]
 RN SEQUENCE OF 1440-1464 FROM N.A.
 RP MEDLINE=90110490; PubMed=2295701;
 RX Willing M.C., Cohn D.H., Byers P.H.;
 RA "Frameshift mutation near the 3' end of the COL1A1 gene of type I
 RT collagen predicts an elongated pro alpha 1(I) chain and results in
 RT osteogenesis imperfecta type I.";
 RL J. Clin. Invest. 85:282-290(1990).
 RN [24]
 RN SEQUENCE OF 1454-1464 FROM N.A.
 RP MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
 RX Maatta A., Bornstein P., Penttinen R.P.;

RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";

Query Match 59.6%; Score 2765; DB 1; Length 1464;
Best Local Similarity 55.4%; Pred. No. 8.2e-103;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGSGSGFGADGAGVAGPKGAGGSGGPGAGPKGSGEA--- 57
Db 179 GPMGSGRGLPGRGSGSGFGADGAGVAGPKGAGGSGGPGAGPKGSGEA--- 238

QY 58 GRPGEAGLPP---GAKGLTSGSGSGP-----DGKTGPPGAGQDGRGPPGPGCAR 105
Db 239 GRPGEAGLPP---GAKGLTSGSGSGP-----DGKTGPPGAGQDGRGPPGPGCAR 298

QY 106 GQAGVMGPPGPKGAAGGPKGAGER-----GVPPGPGAVGPA-----GKDGGAAG 150
Db 299 GQAGVMGPPGPKGAAGGPKGAGER-----GVPPGPGAVGPA-----GKDGGAAG 358

QY 151 GPPGAGPAGGREGGPP-----GSPGPA-----GPGPT 213
Db 359 GPRSEGGVGRGPPGPPGPPGAGAGPAGNPGADQPGKAGGANGAPGAGPFGCARGPS 418

QY 181 GPPGAGPAGGREGGPP-----GSPGPA-----GPGPT 213
Db 419 GPPGAGPAGGREGGPP-----GSPGPA-----GPGPT 478

QY 214 GLPFPGRGSGRGGFGGAGVAGPKGAGGSGGPGAGPKGSGFGGAGRFGGAGLPGAK 273
Db 479 GLPFPGRGSGRGGFGGAGVAGPKGAGGSGGPGAGPKGSGFGGAGRFGGAGLPGAK 538

QY 274 GLTSGPSGPDGKTGPPGAGQDGRGPPGPGARGQAGVMGPPGPKGAAGGPKGAGER 333
Db 539 GLTSGPSGPDGKTGPPGAGQDGRGPPGPGARGQAGVMGPPGPKGAAGGPKGAGER 598

QY 334 GVPGPAGVAGPAGKDGAGAGGPPGAGPAGGSGGPGAGGSGGPGAGGSGGPGAGG 393
Db 599 GVPGPAGVAGPAGKDGAGAGGPPGAGPAGGSGGPGAGGSGGPGAGGSGGPGAGG 658

QY 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
Db 659 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 718

QY 424 GERGGPSGSGFGAGVAGPKGAGGSGGPGAGPKGSGFG-----AGRPGGAGL 474
Db 719 GSQAGPLQGMGPGERGAAGLPGPKDGRDAGPKGADGSPGKDGVRGLTGPFGPGAGAP 778

QY 475 GAK-----GLTSGPSGPDGKTGPPGAGQDGRGPPGARGQ----- 515
Db 779 GAKESGSGGAGTGGARGAGPDRGEPGPPGAGPAGGPPGADGPKGAGGPGDAGAKGDA 838

QY 516 -----AGVMGPPGPKGAAGGPKGAGERGVPPGPA-----VGPAGKDGAGAGAGGPP 561
Db 839 GPPGAGPAGGPPGPGVAGPKGAGGSGGPGAGGSGGPGAGGSGGPGAGGSGGPP 898

QY 562 GPAGP-----AGRGEGGPGSGFGGAGVAGPKGAGGSGGPGAGGSGGPGAGGSGG 618
Db 899 GPAGGEGGPGSGFGGAGVAGPKGAGGSGGPGAGGSGGPGAGGSGGPGAGGSGG 958

QY 619 GPTGLPGRGSGRGGSGFGGAGVAGPKGAGGSGGPGAGGSGGPGAGGSGGPGAGG 675
Db 959 GVVGLPGRGSGRGGSGFGGAGVAGPKGAGGSGGPGAGGSGGPGAGGSGGPGAGG 1018

QY 676 GLPAGKGLTSGPSGPPGDKTGPAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGEP 735
Db 1019 GAGGSGRDSGPKAGDGRGTPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1078

QY 736 GKAGRGVPPGPGVAGPAGKAGE-----AGAGGPPG-----AGPA 771
Db 1079 GPAGARGPAGPQGRGDKGEGTGGGDRGKGRGSGGLGPPGPGSGPQSGGSGAGS 1138

QY 772 GERGGGAGGSP-----GFGGLPAGPAGGAGKPGGQGVGDLGAGPAGGSGAG 820

Db 1139 GPRPPGSGAGAGKGLGLGPGPIGPRGRTGAGVGPAGGPPGPPGPPG 1190

RESULT 2
O76045 PRELIMINARY; PRT; 1461 AA.

AC O76045;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-2004 (TREMBlrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korkko J., Ala-Korkko L., De Paeppe A., Nuytinck L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
RT Identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
RN [7]
RP SEQUENCE FROM N.A.
RA Korkko J.M., Earley J.J., Nuytinck L., DePaeppe A., Prockop D.J.,
RA Ala-Korkko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AAB94054.2;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.

Qy	772	GERGBOGPAGS	---GFGQLPGAPGPGAGKPGECVQVDDILGAPGSPGAG	820
Db	1139	GPRPPGCSAGAPKOGKGLGPGIGPPGRGRTGDTAGPVGPPGPPGPPG	1190	
RESULT 4				
CALL	CANFA	STANDARD;	PRT;	1460 AA.
ID	CALL_CANFA	STANDARD;	PRT;	1460 AA.
AC	QXKSJ7;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	Collagen alpha 1(I) chain precursor.			
GN	NamesCOL1A1;			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_TaxID=9615;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT OI ALA-208.			
RC	TISSUE=Skin;			
RC	MEDLINE=21023337; PubMed=11147834; DOI=10.1006/abbi.2000.2099;			
RA	Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;			
RT	"Sequence of normal canine COL1A1 cDNA and identification of a			
RT	heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case			
RT	canine osteogenesis imperfecta.";			
RL	Arch. Biochem. Biophys. 384:37-46(2000).			
CC	-I- FUNCTION: Type I collagen is a member of group I collagen			
CC	(fibrillar forming collagen).			
CC	-I- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.			
CC	-I- PTM: Prolines at the third position of the tripeptide repeating			
CC	unit (G-X-Y) are hydroxylated in some or all of the chains.			
CC	-I- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfe			
CC	(OI).			
CC	-I- SIMILARITY: Belongs to the fibrillar collagen family.			
CC	-I- SIMILARITY: Contains 1 VWFC domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a colla			
CC	between the Swiss Institute of Bioinformatics and the EMBL outc			
CC	the European Bioinformatics Institute. There are no restrictions			
CC	use by non-profit institutions as long as its content is in			
CC	modified and this statement is not removed. Usage by and for co			
CC	entities requires a license agreement (See http://www.isb-sib.ch/afa			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF153062; AAD34619.1; -			
DR	InterPro; IPR008161; Clg helix.			
DR	InterPro; IPR008160; Collagen.			
DR	InterPro; IPR000885; Fib collagen_C.			
DR	InterPro; IPR001007; VWFC_C.			
DR	Pfam; PF01410; COLFI; 1.			
DR	Pfam; PF01391; Collagen; 18.			
DR	ProDom; PD000007; Clg helix; 2.			
DR	ProDom; PD002078; Fib collagen_C; 1.			
DR	PROSITE; PS01208; VWFC_1; 1.			
DR	PROSITE; PSS0184; VWFC_2; 1.			
KW	Collagen; Disease mutation; Extracellular matrix; Glycoprotein;			
KW	Hydroxylation; Pyrrolidone carboxylic acid; Repeat; Signal;			
KW	Structural protein.			
FT	SIGNAL 1 22			By similarity.
FT	PROPEP 23 157			N-terminal propeptide.
FT	CHAIN 158 1214			Collagen alpha 1(I) chain.
FT	PROPEP 1215 1460			C-terminal propeptide.
FT	DOMAIN 34 92			VWFC.
FT	DOMAIN 158 174			Nonhelical region (N-terminal).
FT	DOMAIN 175 1188			Triple-helical region.
FT	DOMAIN 1189 1214			Nonhelical region (C-terminal).
FT	SITE 741 743			Cell attachment site (Potential).
FT	SITE 1089 1091			Cell attachment site (Potential).
FT	MOD_RES 158 158			Pyrrolidone carboxylic acid (By similarity).
FT	MOD_RES 166 166			Allysine (By similarity)


```
FT MOD_RES 261 261 5-hydroxylysine (By similarity).
FT MOD_RES 1160 1160 3-hydroxyproline (By similarity).
FT CARBOHYD 261 261 O-linked (Gal... ) (By similarity).
FT CARBOHYD 1361 1361 N-linked (GlcNAc... ) (By similarity).
FT VARIANT 208 208 G -> A (in O1; severe).
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 59.5%; Score 2759; DB 1; Length 1460;
Best Local Similarity 55.9%; Pred. No. 1.4e-102;
Matches 559; Conservative 31; Mismatches 230; Indels 180; Gaps 19;

QY 1 GPGEGPTGLPGPPGERGPGS-----RGPPCA-----DGVAGPKPAGERSGPFA 48
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
187 GPGAGCPQGFQPPGEPGASGPMGRPPGPKNGKDDGEAKGRPERGPPGQ 246
QY 49 GPKGSFGEAGRP-----GEAGLPKAGKLTGSPGSPDPDKTTPPGAGOD 93
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
247 GARGLPCTAGLPGMKHGRGFLDGAKDAGPAGPKGEGSPGNGAPQMGRLPGER 306
QY 94 GRPGPPGPPGARGQ---AGVMGPPGPKGAAGE---FGKAGERGVPPGCAVGPAGKGEA 147
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
307 GRFGAGCPAGARGNDGATGAAGPPGTPGAGPPGFCAGVCAKGEAGPQARGSEGPQYR 366
QY 148 GAQGPFGPAGPAGERGEQCP-----AGSPGFQGLPGPAGPPGGEAKGPEQ 192
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 GEPGPPGAGAGPAGNPGADQPGKAGKANGAPGIAAGFPFAGPFGSPGQSPGPK 426
QY 193 GVPGLDGLAP-----GPSGPA-----GEPQPTGLPDPGERGPP 225
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
427 GNSGEPGAPGNKDTGAKGPPGTGQPPGAGEGKRGAGEPQGTGLPFPGERGPP 486
QY 226 GSRGFPAGDGVAGPKPAGERGSPGAGPKGSPGAEAGRPCEAGLPKAGKLTGSPGSPGD 285
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
487 GSRGFPAGDGVAGPKPAGERGSPGAGPKGSPGAEAGRPCEAGLPKAGKLTGSPGSPGD 546
QY 286 GKTGPPGAGDGRCPGPPGARGQAGVMGPPGPKGAAGEPKKAGERGVPPGCAVGA 345
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
547 GKTGPPGAGDGRCPGPPGARGQAGVMGPPGPKGAAGEPKKAGERGVPPGCAVGA 606
QY 346 GKDEAGAGQPPGAPGAGERGQGPAGSPGQGLPGPAGPPGCEAGKPGEQVPGDLGAP 405
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
607 GKDEAGAGQPPGAPGAGERGQGPAGSPGQGLPGPAGPPGCEAGKPGEQVPGDLGAP 666
QY 406 GPSGPAGE-----PGP-----TGLPDPGPPGERGPPGSRGPP 435
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
667 GPSGARGERGFPGERGVQGPFGPAGPRGANGPNDGAKGDAGAPGAPGSGQAGLQGM 726
QY 436 GADGVAGPKPAGERGSPGAPGKSPGE-----AGRPGEAGLPKAK----- 477
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
727 GERGAAGLPKPKDGRDAGPKGADGSPGKDGVRGLTGPFGPPGAPAGDKGEAGSPGA 786
QY 478 GLTSGSPGPDGKTGPPGAGQDGRPPGPPPGARGQ-----AGVMGPP 522
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
787 GPTGARGAPGDRGEPGFPAGFAGPPGADGQPGAKGEPGDAGKGDAGDAGPPGAPTGP 846
QY 523 GPKGAAGEPKKAGERGVPPGPA-----VGPAGKDGAGQAQPPGAPG----- 570
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
847 GPTGNVAGPCKGARSAGPPGATGFPGAAGRVGPPGSGNAGPFPPPGPAKGGKGAR 906
QY 571 GEOGPAGSPGQGLPGPAGPCEAGKPGQGVPGDLGAPGSPGAGEPPTGLPGPGER 630
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
907 GETGPAGRGVGVPPGPPGAGEKSPGADGAGAPGTPGPGIAGQRVVGLPQORGER 966
QY 631 GQPSRGFTGADVAGPKPAGERGSPGAPG---KGSFGEAGRPCEAGLPKAGKLTGSP 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
967 GFFPLPGSPGEPKQKQGSPTSGERGPPGPMGPPGLAGPPGESRGSPGAEKSGPRDGS 1026
QY 688 GSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPPGAGAGEPKKAGERGVPP 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1027 GPKGRDGETGAPGPPGAPGAPGAPGVPAGKNGDRGETGTPAGPAGPIGPVARGAPG 1086
QY 748 GAVGPAKDKGE-----ACAQGPFGP-----AGPAGERGEQCPAGSP 783
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 1087 GPRGDKGTGEGQDRGIKGRHGFSGLOGPPGPPSGFSGSAGSPAGPPGSGAGSP 1146
QY 784 ---GFQGLPGPAGPPGEGAKGPEQGVPGDLGAPGSPGAG 820
Db 1147 GKDGLNGLPPIGPPGPRGRTGADGVPGPVPPGPPGPPG 1186

RESULT 5
Q63079 PRELIMINARY; PRT; 1453 AA.
AC Q63079;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Collagen alpha1 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;
RX MEDLINE=99163824; PubMed=10065941;
RA Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;
RT "Expression of collagen alpha1(I) mRNA variants during tooth and bone
formation in the rat.";
RL J. Dent. Res. 78:11-19(1999).
DR EMBL; Z78279; CAB01633.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR010007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clq helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C.1; 1.
DR PROSITE; PS0184; VWF_C.2; 1.
KW Collagen.
FT NON_TER
SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;

Query Match 59.2%; Score 2748.5; DB 2; Length 1453;
Best Local Similarity 52.3%; Pred. No. 3.7e-102;
Matches 556; Conservative 33; Mismatches 231; Indels 243; Gaps 19;

QY 1 GPPCE---RGPTGLPGPPGERGPGSGRFPG-----ADGVAGPKPAGER 42
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 GPPQDGLPGQGLPGPPGPPGPPGPPGLGLGNFASQMSYGVDEKSAGSVSPGPMGSPGR 176
QY 43 GSPGPAGPKGSPGAGPGEAGLPKAGKLTGSPGSPGDKTGPDPGAGQDGRP---GPP 99
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
177 GLPFPAGAPQGFQGPFGEPGEGGSGPMGPPGPPGPKNGDDGEGAKGKPRGEGRP 236
QY 100 GPPGARGQAGVMGPP-----GPKGAAGBPGK-----A 126
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 GPGARGLPGLTAGLPGMKHGRGFLDGAKGDTGAPGPKGPPGSPGNGTGPQMGPRGLP 296
QY 127 GERGVGPPGCAVGPAGKDGAGAGPPGP-----AGPAGERGEQCP- 167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 GERGRPPGPTAGARGNDGAVGAAGPPGTPGTPGFPGGAAGKAGGAGPQCARGSEGP 356
QY 168 -----AGSPGQGLPGPAGPPGEGACKP 189
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 GVRGPPGPPAGAAGPAGNPGADGQPGKANGANGAPGAGPFGARGPFGPQPSGAP 416
```


DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWC.1; 1.
 DR PROSITE; PS0184; VWC.2; 1.
 KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;
 KW Pyridone carboxylic acid; Repeat; Signal; Structural protein.
 FT SIGNAL 1 22
 FT PROPEP 23 151
 FT CHAIN 152 1207
 FT PROPEP 1208 1453
 FT DOMAIN 29 87
 FT DOMAIN 152 167
 FT DOMAIN 168 181
 FT DOMAIN 182 1207
 FT MOD_RES 152 152
 FT MOD_RES 160 160
 FT MOD_RES 254 254
 FT MOD_RES 1153 1153
 FT CARBOHYD 56 56
 FT CARBOHYD 254 254
 FT CARBOHYD 1354 1354
 FT SITE 734 736
 FT SITE 1082 1084
 FT CONFLICT 1450 1450
 FT SEQUENCE 1453 AA; 137944 MW; 3802E535DF81808 CRC64;
 Query Match 58.9%; Score 2732.5; DB 1; Length 1453;
 Best Local Similarity 54.7%; Pred. No. 1.6e-101;
 Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;
 1 GPCGPGTGLPGPGR-----RGSGRFPAGDAGVAGKPGAGRGSPGPA 48
 180 GPPGAPGQGGPFGPFGGSGMGRGPPGPKGNDGAGKGRGPRGPPGQ 239
 49 GPKGSPGEAGRP-----GEAGLPGKGLTSGSGSPGDKTGPAGQD 93
 240 GARGLPTAGLPGMKHGRFSGLDGKADGAPAGKPGSPGNGAPQGRGLPGR 299
 94 GRPGPPGARGAGVGMVFPKGAAGE-----PKAGRGVFPFGVAVGAGKDGGA 147
 300 GREGPPGTAGAGNDGAVGAAGPPGTGTPGPFPGVAVGAKGEAGRGSEGGVYR 359
 148 GAQGPFGPAGPAGRGQGP-----AGSGFQGLPGAPGAPGAGKPGQ 192
 360 GEPGPPGAGAGPAGNPGADGQKAGANGAPGAGPFGPAGSPGSPGPPGPK 419
 193 GVPFDLGA-----GPSGA-----GEPPTGLPGPBERGGP 225
 420 GNSGEPGAPKGDGTGAKGPGATGVQPPGPGAGEGKRGARCEPFGSLPFGPBERGGP 479
 226 GSRGFPAGDAGVAGKPGAGRGSPGAPKPGSPGAGPGEAGLPGKGLTSGSPGPD 285
 480 GSRGFPAGDAGVAGKPGSPGAGPAGPAGKPGSPGAGRPGEAGLPGKGLTSGSPGPD 539
 286 GKTGPPGAGDQGRPGPPGARGAGVGMVFPKGAAGEPAGRGVAGPAGPAGVGA 345
 540 GKTGPPGAGDQGRPGPPGARGAGVGMVFPKGAAGEPAGRGVAGPAGPAGVGA 599
 346 GKDGGAAGPPGAPGAGRGVAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 405
 600 GKDGGAAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 659
 406 GPSGPAGE-----PGP-----TGLPGPBERGGPGRGPP 435
 660 GPSGARGRGFPGERGVQGPFGPAGPRGNNGAPGNDGAKGDTGAPGAGSQAGLQMP 719
 436 GADGAVAGKPGAGRGSPGAPGKSPG-----AGRPGEAGLPGKGLTSGSP 486
 720 GERGAAGLPGKGRDAGPKGADSGPKDARGLTGPIGPPGAPGADGKAGPAGSP 779
 487 GPD-----GKTGPPGAGDQGRPGPPGARGQ-----AGVMGFP 522
 780 GPTGARGAGPDRGEAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 839

QY 523 GPKGAAGEPKKAGRGVPGPPGA-----VGPAGKDGAGAGQPPGAPG---AGER 570
 DB 840 GFIGNVAGPAGPKGPRGAAGPAGTGPAGAGRVGPPGSGNAGPPPPGFVKEGKGR 899
 QY 571 GQGGPAGSPGFOGLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 630
 DB 900 GTGTGAPRGVEVPPGPPGAPGKSPGADGAPGSPGFTGPGGAGVGVGLPGQGR 959
 QY 631 GQSGRGPFAGDAGVAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 687
 DB 960 GFFGLPGSPGEGKQKPSGSGRGPPGPMGPPGLAGPFGSGRGSGGAGSPGPDGAP 1019
 QY 688 GSPGPDGKTGPPGAGQGRPGPPGARGAGVGMVFPKGAAGEPAGPAGPAGPAGPAG 746
 DB 1020 GAKGDRGETGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1079
 QY 747 -----PGAVGAPKDGAGAGQPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 783
 DB 1080 GPRGDKGTGQDGRGKXHGFGSLQPPGSPGSGPQGGSGAGSPGAPGPPGSGSP 1139
 QY 784 ---GFOGLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 820
 DB 1140 GKGGLGLPGTGPGRGTGDSGAPGPPGPPGPPGPPG 1179
 RESULT 7
 Q810J9 PRELIMINARY; PRT; 1453 AA.
 AC Q810J9
 ID Q810J9
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Procollagen, type I, alpha 1.
 GN Name=Colla1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klapper
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Scheetz T.E.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smaluk U., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX Strausberg R.;
 RA Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
 DB EMBL; BC050014; AAH50014.1;
 DB MGI; MGI:88467; Colla1.
 DB GO; GO:0005615; C:extracellular space; TAS.
 DB InterPro; IPR008161; Clg helix.
 DB InterPro; IPR008160; Collagen.


```
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; C1g_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;

Query Match 58.9%; Score 2732.5; DB 2; Length 1453;
Best Local Similarity 54.7%; Pred. No. 1.6e-101;
Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;

QY 1 GPPGPGTGLPGPGE-----RGPGSRGPPGADGVAGPKGAPAGERSPGPA 48
DB 180 GPPGAPGQGGPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 239
QY 49 GPKGSPGAGRP-----GSAGLPKAGLTGSPGSPGPKTKGPPGPAQD 93
DB 240 GARGLPGTAGLPGMKHKGFGSLDGAKGDAGPAGPKGPPGPPGPPGPPGPPG 299
QY 94 CRPGPPGPPGARGQAGVMGPPGPKAAGE-----PGKAGRGVPPGPPGAVGPAKG 147
DB 300 GRPGPPGTAGRGNDGAVGAGPPTGTPPPGPPGAVGAKGAGPQARGSEPGVR 359
QY 148 GAOGPPGAPGAGERGEOGP-----AGSPGQGLPGPAGPPGEEAKGQEG 192
DB 360 GEPGPPGAGAGPAGNPGADQCGKANGAPGIAGAPFGPARGSPGPPGPPGPK 419
QY 193 GVPGLGAP-----GSPGA-----GEPGTLGPPGPPGPPGPPGPPGPP 225
DB 420 GNSGPPGPNKGDGTAKGEPGATGVQPPGPPGAGEGKRGKAGPGLPGPPGPPG 479
QY 226 SRRGPPGADGVAGPKGAPAGERGSPGAPKSPGAGRPGEAGLPKAGLTGSPGSP 285
DB 480 SRRGPPGADGVAGPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 539
QY 286 GKTGPPGAGQDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 345
DB 540 GKTGPPGAGQDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 599
QY 346 GKDGAAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 405
DB 600 GKGGAAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 659
QY 406 GSPGPPG-----PGP-----TGLPPPGPPGPPGPPGPPGPPGPPGPP 435
DB 660 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 719
QY 436 GADGVAGPKGAPAGERGSPGAPKSPG-----AGRPGEAGLPKAGLTGSPGSP 486
DB 720 GERGAAGLPKGDGDRGADGPKGADGSPKDGARGLTGPIGPPGPPGAGAPGDK 779
QY 487 GPD-----GKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 522
DB 780 GPTGARGAGPDRGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 839
QY 523 GPKGAAGPPKAGRGVPPGPPG-----VGPAGKDGAGAGPPGPPGPPGPPGPP 570
DB 840 GPTGNVAGPPKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 899
QY 571 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 630
DB 900 GTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 959
QY 631 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 687
```

```
DB 960 GFGLPGSPGPKQSPGSSGSRGPPGPPMPPGLAGPPGSGREGSPGABSGPDRGAP 1019
QY 688 GSPGPDGKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 746
DB 1020 GAKGDRGETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1079
QY 747 -----PGAVGPPAGKDCGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPP 783
DB 1080 GPRDKGETGQDGRGKKGHRGFGLOGPPGPPGPPGPPGPPGPPGPPGPPGPP 1139
QY 784 ---GFQGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820
DB 1140 GKGLGGLPGPIGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1179

RESULT 8
Q6LAN8 PRELIMINARY; PRT; 1069 AA.
ID Q6LAN8
AC Q6LAN8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Collagen type I alpha 1 (fragment).
GN Name-COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9714192; PubMed=8988177;
RA Simon M., Peduteou F., O'Brien K., Sirvent N., Kedra D., Guilbaud C.,
RA Turc-Carel C., Dumanski J.P.;
RT "Regulation of the platelet-derived growth factor B-chain gene via
RT fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans
RT and giant-cell fibroblastoma.";
RL Nat. Genet. 15:95-98(1997).
RP SEQUENCE FROM N.A.
RA O'Brien K.P.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X98705; CAA67261.1; -.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0006817; P-phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 16.
DR ProDom; PD000007; C1g_helix; 3.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
FT NON TER 1069 1069
SQ SEQUENCE 1069 AA; 97445 MW; BE279B10572FB980 CRC64;

Query Match 57.8%; Score 2681.5; DB 2; Length 1069;
Best Local Similarity 58.4%; Pred. No. 1.4e-99;
Matches 524; Conservative 28; Mismatches 229; Indels 117; Gaps 10;

QY 1 GPPGPGTGLPGPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 60
DB 203 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 262
QY 61 GEAGLPKAGLTGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 120
DB 263 GMKGRGFGSLDGAKGDAGPAGPKGPPGPPGPPGPPGPPGPPGPPGPPGPPG 322
QY 121 GEPKAGRGVPPGPPGAVP-----AGKDGAGAGQPPGPPGPPGPPGPPGPP 167
DB 323 GNDGATGAAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 382
```



```

QY 168 -----AGSPGFGQLPGPAGPGEAGKPGEGVPCDLGAP--- 201
Db 383 GPAGNPGADGQCAKANGAPGAGPFGARGSGGQGGPGPKNGSGEFGAGPSK 442
QY 202 -----GPSGA-----GPPGPTGLPGPPGEGGSGRGGFGPCADGVA 237
Db 443 GDTGAKGEPGVGVGPPGPGAGEGKRGARGPPTGLPGPPGEGGSGRGGFGPCADGVA 502
QY 238 GPKGAGERSGPGAGPKSGPCEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPGAGOD 297
Db 503 GPKGAGERSGPGAGPKSGPCEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPGAGOD 562
QY 298 GPPGPPPCARGQAGVWGFPGPKGAAGEPKKAGRGVPPGPGAVGPGAGKDGGAAGQGP 357
Db 563 GPPGPPPCARGQAGVWGFPGPKGAAGEPKKAGRGVPPGPGAVGPGAGKDGGAAGQGP 622
QY 358 GPAGPAGERGEGGPGAGSGFGQGLPGPAGPGEAGKPGEGVPCDLGAPGSGPAGERGPT 417
Db 623 GPAGPAGERGEGGPGAGSGFGQGLPGPAGPGEAGKPGEGVPCDLGAPGSGPAGERG 678
QY 418 GLPGPPGRRGGSGRGGFGPCADGVAAGPKGAGERSGPGAGPKGSPGAGRPGEAGLPGAK 477
Db 679 -----RGPPGRRGVGQPPGPGAGPAGANGAPGNDGAKGAGPAGPSGQAP 724
QY 478 GLTSGSPGSGDGTGTPGPGAGODRPPGPPGARGOAGVMGPPGKGAAGSPGKAGER 537
Db 725 GLQWMP-----GERGAAGLPGKGDGRDA-----GPKGADGSGPKDGVR 763
QY 538 GVPGPAGVAGPKDGEAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 591
Db 764 GLTGPFGPPGAGPAGDKGSGSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 823
QY 592 GEAGKPGEGVPCDLGAPGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 645
Db 824 GAKGEPDAGAKGADGAGPFPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 883
QY 646 GPKGAGERSGPGAGPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPGAGOD 705
Db 884 GPPGSGNAGPFPAGPAGKGGKPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 943
QY 706 GRPFPGPAGARGQAGVWGFPGPKGAAGEPKKAGRGVPPGPGAVGPGAGKDGGAAGQGP 765
Db 944 GAPGTPGQGTAGQGVVGLPGQRRGRRGFPGLPGSPGEGKQGPSGASGERGPPGMPGP 1003
QY 766 GPAGPAGERGEO---GPAGSPGQGLPGPAGPGEAGKPGEGVPCDLGAPGSGPAG 820
Db 1004 GLAGPPGSRGREGAPGAEPSGRDGSFGAKGDRGETGPAGPAGPAGPAGPAGPAG 1061

```

RESULT 9

```

CALL CHICK
ID CALL CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-153 FROM N.A.
RA MEDLINE=88056316; PubMed=3678834; DOI=10.1016/0378-1119(87)90159-4;
RX Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RL Gene 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.

```

```

RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
RL first intron of the chicken pro-alpha 1(I) collagen gene.";
RN J. Biol. Chem. 262:13323-13332(1987).
RP SEQUENCE OF 152-1187.
RX MEDLINE=82231395; PubMed=7093229;
RA Hightberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H.,
RX Gross J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
RL complete primary structure of the helical portion of the chick skin
RN collagen alpha 1(I) chain.";
RP Biochemistry 21:2048-2055(1982).
RX SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glimcher M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
RL of the alpha 1 chain of chicken bone collagen.";
RN Biochem. Biophys. Res. Commun. 48:720-726(1972).
RX SEQUENCE OF 981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RL alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RN including the carboxy-terminal propeptide sequences.";
RX Biochemistry 20:996-1006(1981).
RX SEQUENCE OF 1311-1453 FROM N.A.
RX MEDLINE=80134546; PubMed=6987088; DOI=10.1016/0014-5793(80)80761-7;
RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
RL carboxyl end of pro alpha 1(I)-chains.";
RX FEBS Lett. 111:61-65(1980).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains. Pro-1153 is the only 3-hydroxypro and the only
CC hydroxylated proline in position X.
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 VWF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M17839; AAA48704.1; -.
CC EMBL; M17838; AAA48704.1; JOINED.
CC EMBL; V00401; CAA23695.1; -.
CC EMBL; M10571; AAA48671.1; AUT_SEQ.
CC EMBL; M17607; AAA48672.1; -.
CC PIR; A27179; A27179.
CC PIR; I50629; I50629.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_Collagen_C.
CC InterPro; IPR010007; VWF_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC Pfam; PF00093; VWC; 1.
CC ProDom; PD000007; Clg_helix; 2.

```



```
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen; Direct protein sequencing; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;
KW Signal; Structural protein.
FT SIGNAL 1 22
FT PROPEP 23 151
FT CHAIN 152 1205
FT PROPEP 1206 1453
FT DOMAIN 31 89
FT MOD_RES 152 152
FT MOD_RES 160 160
FT MOD_RES 254 254
FT MOD_RES 851 851
FT MOD_RES 1081 1081
FT MOD_RES 1097 1097
FT MOD_RES 1153 1153
FT CARBOHYD 254 254
FT CARBOHYD 1354 1354
FT CONFLICT 1187 1187
FT CONFLICT 1441 1441
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Query Match 57.7%; Score 2679; DB 1; Length 1453;
Best Local Similarity 53.7%; Pred. No. 2.1e-99;
Matches 540; Conservative 30; Mismatches 250; Indels 186; Gaps 16;

QY 1 GPPGPGTGLPCPPGRRGSGRFFPGADGVAGPKGPA-----GER 42
DB 174 GPRGLGPPGAPGQFGQPGGEPGSGASGPMGPRGPPGPKNGDDGEAGKPRQGR 233
QY 43 GSPGAPGKSGEAGRPCEAGLPCAKGLTSGSPGPDGKTGPPGAGOD----- 93
DB 234 GPPGQAGRLPDTAGLPGMKHRSFGLDGAQKQPGAGPKGFSGNGENAPQCMGR 293
QY 94 -----GRPGPGPPGARGQ-----AGVMGPPGKGAAGE-----PGKAGERGVPPGAVGPA 141
DB 294 GLPGERGRPPGSPAGARGNDGAPCAAGPPGTPGAGPPGPPGGAAGKGTGPGARGSE 353
QY 142 GHDGAGAGPPGPPGAPAGERGEQGP-----AGSGFQGLPAPGPPGEA 186
DB 354 GPQSGRGEPPGPPGAGAGPAGNPGADGQPGAKGATGAPGIAGAPGFFGARGSPGQPS 413
QY 187 GRPGQGVPGDLGAP-----GPSGPA-----GEPGPTGLPGPP 219
DB 414 GAPGPKGNSGEPGAPGNKGDGAKGPPGAGVQGPFGAGEGKKGARGEPGAGLPGA 473
QY 220 GERGGFSGRFPFGADGVAGPKGAPAGERGSPGAPGKSPGGEAGRPCEAGLPCAKGLTSP 279
DB 474 GERGAPSGRFPFGADGIAGPKGPPGERSPGAVGPKSPGGEAGRPCEAGLPCAKGLTSP 533
QY 280 GSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPKAGRGVPPGP 339
DB 534 GSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPKGPGERGAPGP 593
QY 340 GAVGAPGKDGAGAGQPPGAPGAPGERGPPGQAGSPGFGQLPAPGCEAGKPGEQVP 399
DB 594 GAVGAPGKDGAGAGQPPGTPGAGERGQGPAGAPGFGQLPAPGCEAGKPGEQVP 653
QY 400 GDLGAPGSPSPAGE-----PGP-----TCLPFPGERGPP 429
DB 654 GNAGAPGAPAGERGPPGGERGVQGPFGPPGPRGANGAPNDGAKDAGAPAGNEGPP 713
QY 430 GSRGFPFGADGVAGPKGAPAGERGSPG-----AGPKGSPGEAGRP-----GEA 471
DB 714 GLECHPMGERGAAGLPCAKGDRGDPGPKGADGAPKDGRLGLTGPAGPAGAPDKGEA 773
QY 472 GLPGAAGLTSFGSPGPDGKTGPPGPA-----QDGRPPGP 507
DB 774 GPPGAPGTPGARGAPGDRGEPGPPGAPGAPGPPGADGQPGAKGETGDAGAKGDAGPPGPA 833

QY 508 GPPGARGQAGVMGPPGPKGAAGEPGKAGERGVPPGPPGAVGAPGKDGKGEAGAGQPPGPGAPA 567
DB 834 GPTGAPGAGVAGPAGPKARGSGAGPPGATGPPCAAGRGVGGPSGNTGLGPPGPGAGKZ 893
QY 568 GE---RGSGQAGSPGFGQLPAGPAGPCEAGKPGQGVPGDLGAPGSPGAPGPTGLP 624
DB 894 GSKPRGETGAPRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 953
QY 625 GPPGRRGSGRFPFGADGVAGPKGAPAGERGSPGAPG---KSGFGEAGRFEAGLPGAK 681
DB 954 GQGERGFPGLPGPSGPRGKQPSGASGERGPPGPMGPPGLAGPPEAGREGAPGAEGAP 1013
QY 682 GLTSGPSGPPGDKTGTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPKKAGER 741
DB 1014 GRDGAAGPKGRDGETGAPGPPGAPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1073
QY 742 GVPGPPGAVGAPGKDG-----AGACQGPP-----GPAKAGER 774
DB 1074 GPAGPQGRGDKGTGEGQDGRGMKGRGFGSLQGPFGPPGAPGEGQPSGASGAPGPRGP 1133
QY 775 GEOQAPGSPGQGLPAGPAGPGEAGKPGEQGVPGDGLGAPGSPGAPG 820
DB 1134 GSAGAGKGLGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1179

RESULT 10
O93251 PRELIMINARY; PRT; 1445 AA.
AC O93251;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
GN Name=alpha 1 type I collagen;
OS Rana catesbeiana (Bull frog);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95294154; PubMed=10367734; DOI=10.1016/S0945-053X(99)00005-0;
RA Ashina K., Utoh R., Ohara M., Yoshizato K.;
RT "Cell-type specific and thyroid hormone-dependent expression of genes
RT of  $\alpha 1(I)$  and  $\alpha 2(I)$  collagen in intestine during
RT amphibian metamorphosis.";
RL Matrix Biol. 18:89-103(1999).
DR EMBL; AB015440; BAA29028.1;
DR GO; GO:0005581; C:cytoplasm; IEA.
DR GO; GO:0005737; P:extracellular matrix structural constituent; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Cig helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWFC_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Cig_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWFC_1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen.
SQ SEQUENCE 1445 AA; 137251 MW; F59BB550C9873F04 CRC64;

Query Match 55.3%; Score 2568; DB 2; Length 1445;
Best Local Similarity 53.6%; Pred. No. 5.4e-95;
Matches 516; Conservative 43; Mismatches 259; Indels 144; Gaps 17;

QY 1 GPPGPGTGLPGGPGERGPGS-----RGFPAGDVAGPKGPA-----GER 42
```


QY 414 -----PGTGLPGP-----PGRGGSRGPP----- 435
Db GERGGQAGAGPRGSGPCSGNDGAKGAGAGAPGGRGPPGQGMFGERGSGAMPKAG 728
QY 436 -----GADGAGVAGKPGAGRGSP-----GPAGPKGSGPGEAGRP 468
Db GDRGDAGTKGADGAPGKDGAGRLTGPIGPPGSGAPGDKGEGSGPAGPTGARGSPGER 788
QY 469 GEAGLPGAKGLTGSPGS---PGPDGKTGPPAGQDGRPPGPPGARGQAGVGMFFGPKP 525
Db GEPGAPGAGICGPPGADGQPGKAGSGDAGPKGDAGAPGAGPTGAPGAGNVGAPGPK 848
QY 526 GAAGPGRGAGRGVPGPCAVGAPGKDGCEAGACQAGPPGAPGAGERGEQAGSPGFGQLP 585
Db GTRGAAGPPGATGPPGAAGRLGPPGSGNVAGPPGPPGCKGKAGSGRGTGAPGRSGEP 908
QY 586 SPAGPP---GEAGKGEQGVCDLGAAGPSPGAPGPTGLGPPGGERGGSGRFFPAD 642
Db GPAGPPGSGEKSFGSDGAPAGIPIPGQIAGQGVVGLFGQGERGFGSLPGPAGEP 968
QY 643 GVAGPKGAPAGRGSGPPAGP-----KGSPPGAGRPGEAGLPGAK-----GLT 684
Db GKQSGSPNGERGPPGSPGPPGLGPPGCEPGRGSGSPGSEAGPGRDGSFGPKGDRGSGPS 1028
QY 685 GSPSGPDGKTGPPGAPGQD---CRPPGPPGARGQAGVGMFFGPKGAGPCKAGE- 740
Db GPPGAPGAPGAPGPPVGPAGKNGDRGETGAPGAPGAPGSGVGRGAPGAPGAGKDGAGEQ 1088
QY 741 -----RGVGPAGVAGPAGKDGCEAGACQAGPPGAPGAGERGEQAGSPGFGQLGPA 792
Db GERGMKGRHNGMGPPGPPGSSGEGQAGPDSGAPGPRGPPGSSGSGTKDGVNGLPGPI 1148
QY 793 GPPGAGRGVGMQPPGPPGSSGEGQAGPDSGAPGPRGPPGSSGSGTKDGVNGLPGPI 820
Db GPPGPRGRNDGVPAGPPGPPGPPG 1176
RESULT 12
Q640B2 PRELIMINARY; PRT; 1449 AA.
AC Q640B2
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RX PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RA Klein S., Gerhard D.S.;
RL Submitted (Sep-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC082718; AAH82718.1, -;
KW Hypothetical protein.
SQ SEQUENCE 1449 AA; 137539 MW; 456639B1687A3B4B CRC64;
Query Match 55.1%; Score 2558.5; DB 2; Length 1449;
Best Local Similarity 53.2%; Pred. No. 1.3e-94;
Matches 508; Conservative 44; Mismatches 268; Indels 135; Gaps 13;
QY 1 GPPGEPGTGLPGRPPGRRGPGS-----RFPAGDVAGPKGPA-----GER 42
Db 222 GKPRPGRGPPPGQARGLPGTAGLPGMKHRRFNGLDGAKGDSGAPGKGEPSGPN 281
QY 43 GSPGAPGKSPGAGRGPEAGLPGAKGLTSGSPSPGPDGKTGPPG-----A 90
Db 282 GAPQVGPRLGPRGRGPPGSPGAGARGNDGAPGAGPNGPTGAGPPGPPGPGVGP 341
QY 91 QDGRPPGPPGARGQAGVGMFFGPKGAGCEPGKAGRGVPPGPGAVGAPGKDGCEAG 150
Db 342 GPQSRGPDGQGRGEPGAPGQAGAGPSNPNPSDQAGAKGATGAPGAGPFGAR 401
QY 151 GPPGAPGAGRGQAGSPGFGQLGPPGAPGPEAGKPGQGVPGDLGAPGSPGAPGE 210
Db 402 GAPGQGGAPGAPGKNNGEPGACQNKGEPCAKGESGAPGVQPPGSGEGKGRSGEP 461
QY 211 GPTGLPGRGPPGSRGPPGAGVAGKPGAPGERSGPPGAPGKSGPGEAGRPGEAGLP 270
Db 462 GPAGPPGAPGARGAPGSRGPPGSDGASGKPGGPPGPPGAGKSGPESGRPGEPLG 521
QY 271 GAKGLTSGSPGPDGKTGPPGAPGQDGRPPGARGQAGVGMFFGPKGAGPCKA 330
Db 522 GAKGLTSGSPGPDGKTGPPGAPGQDGRPPGARGQAGVGMFFGPKGAGPCKA 581
QY 331 GERGVPPGAVGAPGKDGCEAGACQAGPPGAPGAGERGEQAGSPGFGQLGPPGAPG 390
Db 582 GERGVPPGAVGAPGKDGCEAGACQAGPPGAPGAGERGEQAGSPGFGQLGPPGAPG 641
QY 391 GKPGQGVPGDLGAPGSPGAPG-----PGTGLPGRGPPGSGRPPGAGDVAGPKPA 447
Db 642 GKPGQGVPGDVGSPGAPGARGRGFPGERGAQGPFGPQGRGSGNAGPNDGAKGEAG 701
QY 448 GERGSPGAPGKSGSPGAGR-----PGEA 471
Db 702 GAPGGPPGLQGMPPGERGAGGLPGAKGDRDQGVKSGSDGAPGKDGVRGLTGTP 761
QY 472 GLPGAK-----GLTSGSPGPPDGKTGPPGAPGQDGRPPGPPGARGQ----- 515
Db 762 GAPDKGEAGPAGPAGTARGAGPGERGESPPGAPGAPGPPGADGQFGAKGEQDAGAK 821
QY 516 -----AGVGMPPGPKGAAAGEPGKAGRGVPG-----PGAVGAPKDGCEAGACQ 561
Db 822 GDAGPPGAPGATGAPGAPAGALGATGPKGARGAPGAPGATGTRVGPVPPGAPGNSGP 881
QY 562 GPAGPAGE-----RGEQGRAGSPGFGQLGPPGAPGKGEQGVPGDLGAPGSPGA 615
Db 882 GPSGAPGKGEAGKPRGTGTPAGRGPEGPAAGPAGPPGKSGPSGSDGAPAGPIPG 941
QY 616 GEPGPTGLPGRPPGRRGPPGSRGFPAGDVAGKPGAPGERSGPPGAPG-----K 671
Db 942 GSRGTVGLPGRGRRGPPGSLPGVGPCKQKQPSGSGRPPGPPGPPGAGPGEAGRE 1001
QY 672 ---PGEAGLPGAKGLT-----GSPSGPDGKTGPPGAPQDGRPPGPPG 714
Db 1002 GAPGSEGAPGRDGAAGPKDRGEGGAPGPPCAPGAPGAPGVPVGPAGKSGDRGETGAPGA 1061
QY 715 GARGQAGVGMPPGPKGAAAGEPGKAGE-----RGVPPGPPGAVGAPKDGCEAGACQ 765
Db 1062 GPAGVAGARGAPGAPGPRGDKGEAGEQGERGMKGRHNGFNGPAGPPGPPGSSG 1121

766 GPAGPAGERGEQCPAGSPGFQGLPGAPGPPGBAGKPKGVDPDI.GAPEGSGPAG 820
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1122 GPAPRGPPGGSSGNPKGDCGANGLPGIPGPGRGTGDVGPAGPPGPFGPPGPG 1176
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 13

8R02B5 PRELIMINARY; PRT; 1449 AA.

AC Q802B5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Collal-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zesberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
FAhey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences." ;
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[2]

RN [3]

RP Dev. Dyn. 225:384-391(2002).

TISSUE=Embryo;
RC Klein S., Strausberg R.;
RX Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RL EMBL; BC049829; AAH49829.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:citoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000895; Fib_collagen_C.
DR InterPro; IPR009041; PMF_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1_-
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; WVC; 1.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.

RESULT 14
 Q6PCL3 AC Q6PCL3 PRELIMINARY; PRT; 1225 AA.
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Colial protein.
 GN Name=Colial;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Woxley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krywinski M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RA Strauberg R.;
 RL Submitted (OCT-2003) to the ENBL/GenBank/DBJ databases.
 DR EMBL; BC059281; AAH59281.1; -
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; Clq helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008885; Fib_collagen_C.
 DR InterPro; IPR009041; PMP_SGCI.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF00093; VWC; 1.
 DR ProDom; PD000007; Clq_helix; 2.
 DR ProDom; PD02078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C; 1.
 DR PROSITE; PS01084; VWF_C2; 1.
 KW Collagen.
 SQ SEQUENCE 1225 AA; 117860 MW; B6B86CB4457F4D9 CRC64;
 Query Match 53.8%; Score 2494; DB 2; Length 1225;
 Best Local Similarity 56.0%; Pred No. 4.2e-92;
 Matches 510; Conservative 29; Mismatches 227; Indels 144; Gaps 20;
 12 PGPCERGG-----PGSRGFGGADGAGKPGKPGAGERSGPGAGKSGSGEAGRPGCEAGLP 66
 73 PNPQRRECCAFCEPEEVVSNSEVDV-GVEGPKDGPQGGPRGPGVGGPRGDIQGQCLP 131
 67 GAKGLTSGPGSGPDGK-----TGPPGAGQDGRPPGPPGARGQA 108

Db 132 GPPCHPFPFPGLGNGFASQMSYGYDEKSAVSVPFPMGSPGRLGLPGPPGARGPQ---- 188
 Qy 109 GVMGFPKGAAGEPGKAGRGVPGFPGVAGPAGKDGAGAGQGGP-----GPAGPAGERGEQ 165
 Db 189 ---GFQGFPPGEPGEPGSGPMGPRGPPGPGKNGDDGDEAGKPGRPGPPGPGQARGLP 245
 Qy 166 GPAGSP-----GFQGLPGPAGPGEAGKPGEGVPGDLGNAPGSPGAGERGPTGLPGPP 219
 Db 246 GTAGLPGMKHGRGFGSLDGAKGADGAPAGPKGEPGSPGENGAPGQMGPRGLPGERGRPGP 305
 Qy 220 GERGGPSGRFPAGADGVAGPKPGAGERGSGPAGPKGSPGEAGRPGAGLPGAKGLTGPS 279
 Db 306 GT---AGARGNDGAVGAGPGPT-----GPTGPPGFPAGVAGKAGGAGPGAR----- 350
 Qy 280 GSPGFDGKTGPPGAGQDGRPPGPPGARGQAGVMGFPFGKGAAGSPFGKAGRGVPGPP 339
 Db 351 GSEGPQGVRRGPPGPPGAGAAAGPAGNFCADQ-----PGAKGANGAPGIA---GAPGFP 401
 Qy 340 GAVCPAGKDGAGAGAGQPPGAGPAGERGEOGPAGSPFGQGLPGPAGPGEAGKRGEOGVP 399
 Db 402 GARGP-----SGPQSPSGPPGPKNSGE-----PGAPGNKGDGTGAKGEPGATGVQGP 449
 Qy 400 GDLGAPGSPGAPGPPGTLGPPGPRGSGRGGPFGADGVAGPKGAGERGSGPAGPK 459
 Db 450 GPAGEEKRGARGEPGSGLPFGPPEERGGPGSRGFGADGVAGPKGSGERKAGFGPAGPK 509
 Qy 460 GSPGAGRPGGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVM 519
 Db 510 GSPGAGRPGGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVM 569
 Qy 520 GFPQPKAAGEPKKAGRGVPPGAVGAPKDGDEAGAGQPPGAGPAGERGEOGPAGSP 579
 Db 570 GFPQPKTAGEPKAGRGVPPGAVGAPKDGDEAGAGQAGQAGPAGPAGERGEOGPAGSP 629
 Qy 580 GFQGLPGPAGPPGAGKPGQGVPGDLGAPGSPGAGE-----PGP----- 620
 Db 630 GFQGLPGPAGPPGAGKPGQGVPGDLGAPGSPGARGERFPGERGVQGPAGPPAGPRGN 689
 Qy 621 -----TGLPSPGPRGSGRGGFPFGADGVAGPKGPGAGRSGSPGAPKSGPGE- 668
 Db 690 GAPGNDGAKGDTGAPGAPGSGQAGFGLQGMPCERGAAGLPGPKGDRGADGPKGADSGPKD 749
 Qy 669 -----ACRPGGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 720
 Db 750 GARGLTGFIQPPGAPGAGDKGAGSGSPGPGTARGAPGDRGAGPPGAPGAGPPGAP 809
 Qy 721 GVMGFPKGAAGEPGKAGRGVPGVAGVPGKDGDEAGAGQPPGAPGAGERGE----- 776
 Db 810 GAPGAPGVGAPGKNGDRGETGPAGPAGTIGPARGAPGAPGQGRGDKGETGEOGDRIK 869
 Qy 777 -----QGPAGSPGPGQGLPGPA-----GPPGAGKPGGE---QGVPGDLGAPGP- 815
 Db 870 GHRGFGSLQGPSPGSGGSGAGSPAGPRPGPSAGSPGKDLGLGLPGIPGPGPR 929
 Qy 816 -----SGPAG 820
 Db 930 GRTGDSGPAG 939
 RESULT 15
 Q14046 ID Q14046 PRELIMINARY; PRT; 1160 AA.
 AC Q14046;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE COL2A1 protein precursor (fragment).
 GN Name=COL2A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 88.8873 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSECPGVGRGPGPPGAGA.....PPSGDAGPPGPPGAPGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 16Dec04:.*
1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003s:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3070	100.0	544	ADM48393	Adm48393 Recombina
2	2966	96.6	1057	AAy84541	AAy84541 Amino aci
3	2966	96.6	1057	AAy84544	AAy84544 A human c
4	2966	96.6	1058	AAy84403	AAy84403 Amino aci
5	2966	96.6	1107	AAr89472	AAr89472 Collagen/
6	2966	96.6	1107	AAy84540	AAy84540 Amino aci
7	2966	96.6	1161	ADe87050	ADe87050 Human pan
8	2966	96.6	1169	AAr89469	AAr89469 Collagen/
9	2966	96.6	1169	AAy84537	AAy84537 Amino aci
10	2966	96.6	1171	AAr89470	AAr89470 Collagen/
11	2966	96.6	1171	AAy84538	AAy84538 A chimeri
12	2966	96.6	1211	ADe87057	ADe87057 Human pan
13	2966	96.6	1226	ADe87062	ADe87062 Human pan
14	2966	96.6	1388	AAr89471	AAr89471 Collagen/
15	2966	96.6	1411	AAy56800	AAy56800 Human pre
16	2966	96.6	1461	ABg93947	ABg93947 Human pol
17	2966	96.6	1461	AAw68485	AAw68485 Human rec
18	2966	96.6	1464	AAw68485	AAw68485 Human pro
19	2966	96.6	1464	AAU14136	AAU14136 Human nov
20	2966	96.6	1464	AAy50764	AAy50764 Human tum
21	2966	96.6	1464	ABP68610	ABP68610 Human pan
22	2966	96.6	1464	ABU54471	ABU54471 Human tum
23	2966	96.6	1464	ABR47417	ABR47417 Breast ca
24	2966	96.6	1464	ABR92064	ABR92064 Human cer
25	2966	96.6	1464	ADD14142	ADD14142 Human src

26	2966	96.6	1464	7	ADD45059	ADD45059 Human pro
27	2966	96.6	1464	7	ADD45055	ADD45055 Human pro
28	2966	96.6	1464	7	ADD45051	ADD45051 Human pro
29	2966	96.6	1464	7	ADP65246	ADP65246 Human alp
30	2966	96.6	1464	7	ADP65246	ADP65246 Human alp
31	2966	96.6	1464	8	ADQ19470	ADQ19470 Human sof
32	2966	96.6	1464	8	ADQ29653	ADQ29653 Human col
33	2966	96.6	1464	8	ADP16800	ADP16800 Human col
34	2966	96.6	1464	8	ADP16800	ADP16800 Human col
35	2966	96.6	1464	8	ADR99144	ADR99144 Collagen,
36	2958	96.4	1388	3	AAy84539	AAy84539 Amino aci
37	2936	95.6	1463	4	AAE02532	AAE02532 Bovine al
38	2920	95.1	1536	7	ADe87051	ADe87051 Human pan
39	2872	93.6	1453	7	ADD45053	ADD45053 Rat Prote
40	2872	93.6	1453	7	ADD45057	ADD45057 Rat Prote
41	2872	93.6	1453	7	ADD48341	ADD48341 Rat Prote
42	2872	93.6	1453	7	ADD45049	ADD45049 Rat Prote
43	2872	93.6	1453	7	ADD48337	ADD48337 Rat Prote
44	2872	93.6	1453	7	ADD48345	ADD48345 Rat Prote
45	2857	93.1	822	2	AAy06240	AAy06240 Mouse rec

ALIGNMENTS

RESULT 1
ADM48393
ID ADM48393 standard; protein; 544 AA.
XX
AC ADM48393;
XX
DT 03-JUN-2004 (first entry)
XX
DE Recombinant gelatin-like polypeptide Hu-deam.
XX
KW Plasma substitute; Gelatin-like protein; plasma expander.
XX
OS Synthetic.
XX
FN EPI398324-A1.
XX
PD 17-MAR-2004.
XX
PF 11-SEP-2002; 2002EP-00078745.
XX
PR 11-SEP-2002; 2002EP-00078745.
XX
PA (FUJF) FUJI PHOTO FILM BV.
XX
PI Bouwstra JB, Toda Y;
XX
DR WPI, 2004-229415/22.
XX
PT Composition useful as substitute for plasma, comprises solution of saline
PT and recombinant gelatin-like protein having colloid osmotic function.
XX
Example 1; SEQ ID NO 4; 31pp; English.

The present sequence is the protein sequence of recombinant gelatin-like protein Hu-deam. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 ADM48390 or Hu-deam. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing Ige antibodies. The gelatin-like proteins can be covalently attached to pharmaceutical active compounds. After administration, the coupled

CC A fusion protein (AAR89469) comprises the alpha-helical region of human
CC collagen I(a) linked to the human mature bone morphogenetic protein 2B
CC (BMP2B). It can be expressed in *Escherichia coli* transformants carrying a
CC vector incorporating a chimeric gene (AAT16515) coding for the fusion.
CC The BMP moiety induces osteogenesis, while the collagen moiety provides
CC an integral substratum or scaffolding for the BMP and cells involved in
CC reconstruction and growth. The fusion protein provides sustained release
CC and delivery of BMP to a target tissue
XX
SQ Sequence 1169 AA;

Query Match 95.6%; Score 2966; DB 2; Length 1169;
Best Local Similarity 95.6%; Pred. NO. 3.1e-16;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0

QY 1 GSEPGVGVRGEPGPPGAGAAPGADPCADGCEPGAKGADGAPGIAGAPGFGARGPSGPE 60
DB 201 GSEPGVGVRGEPGPPGAGAGPAGNPGADQPGAKGANGAPGIAGAPGFGARGPSGPP 260
QY 61 GPGGPPGPKDSDGSPGAPSGKDTGAKGEPGVGVGPPGAGEGKPGARGEPGPTGLP 120
DB 261 GPGGPPGPKGNSGEPGAPSGKDTGAKGEPGVGVGPPGAGEGKPGARGEPGPTGLP 320
QY 121 GPPERGGPSGRPGADCVAGPKPAGERGSPGAPGKSPGCEAGRCEAGLPGAKGLT 180
DB 321 GPPERGGPSGRPGADGVAGPKPAGERGSPGAPGKSPGCEAGRCEAGLPGAKGLT 380
QY 181 GSPGSPGPDGKTGPPGAGEDGRPGPPGARGGAGVGMGPPGKGAAGEPGKAGERGVP 240
DB 381 GSPGSPGPDGKTGPPGAGQDGRPGPPGARGQGVGMGPPGKGAAGEPGKAGERGVP 440
QY 241 GPPGAVGAPGKDGABAGGPPGPPGAPGAGERGEEGPGAGSPGFEGLFGPPGAGEAGKPGSE 300
DB 441 GPPGAVGAPGKDGABAGAQPPGPPGAPGAGERGEEGPGAGSPGFEGLFGPPGAGEAGKPGSQ 500
QY 301 GVPDGLGAPGSGARGEPGFPGERGVGPPGPPGADGAPDDGAKGDAGAPGAPSE 360
DB 501 GVPDGLGAPGSGARGERGFPGERGVGPPGPPGAPGPRGANGAPGNDGAKGDAGAPGAPSQ 560
QY 361 GAPGLEGMWGERGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGPIGPPGAPGAPGDK 420
DB 561 GAPGLQGMWBERGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGPIGPPGAPGAPGDK 620
QY 421 GESGSPGAPGTGARGAPGDRGEPGPPGAPGAGPFGADGPFKAGEPCDAGAKGDAGPP 480
DB 621 GESGSPGAPGTGARGAPGDRGEPGPPGAPGAGPFGADGPFKAGEPCDAGAKGDAGPP 680
QY 481 GPAGPAGPPGPIGVGAPGAKGARGSAGPPGATGPGAAAGRVGPPGSDGAPGPPGPPA 540
DB 681 GPAGPAGPPGPIGVNVPAGKARGSAGPPGATGPGAAAGRVGPPGSDGAPGPPGPPA 740
QY 541 GKEG 544
DB 741 GKEG 744

RESULT 9
AAY84537
ID AAY84537 standard; protein; 1169 AA.
AC
XX AAY84537;
XX
XX 25-JUL-2000 (first entry)
XX
XX Amino acid sequence of a chimeric collagen 1 (alpha1)/BMP-2B protein.
XX
XX Extracellular matrix protein; self aggregation; hydroxylated proline;
KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW collagen; fibrinogen; fibronectin; post translational hydroxylation;
KW ss. Bone morphogenic protein; BMP-2B; chimera.
XX
XX Homo sapiens.
OS Unidentified.

Db 381 GSPGSPGDKTGPDPGAGQDGRPPGPPGARGQAGVMGPPGKAGGFGKAGRGVP 440
 QY 241 GPPGAVGAGKDGAGAGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 300
 Db 441 GPPGAVGAGKDGAGAGQPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 500
 QY 301 GVPDGLGAPGSGARGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 360
 Db 501 GVPDGLGAPGSGARGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 560
 QY 361 GAPLEGMPGPPGAGAGLPDPKDRDAGPKGADGSPGKGVRLTGPDPGAGAGDK 420
 Db 561 GAPLEGMPGPPGAGAGLPDPKDRDAGPKGADGSPGKGVRLTGPDPGAGAGDK 620
 QY 421 GESGSPGAGTGPAGAPGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 480
 Db 621 GESGSPGAGTGPAGAPGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 680
 QY 481 GPAGPAGPPGPPGAGAGKAGSAGPPGATGPPGAGRVGPPGPPGAGPPGPPGPA 540
 Db 681 GPAGPAGPPGPPGAGAGKAGSAGPPGATGPPGAGRVGPPGPPGAGPPGPPGPA 740
 QY 541 GKGE 544
 Db 741 GKGE 744

RESULT 10
 AAR89470
 ID AAR89470 standard; protein; 1171 AA.
 AC AAR89470;
 DT 01-OCT-1996 (first entry)
 DE Collagen/TGF-beta-1 fusion protein.
 KW Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
 KW bone formation; tissue repair; fusion protein.
 OS Synthetic.

Key Location/Qualifiers
 Domain 1..1057
 /label= Collagen-IA
 /note= "collagen IA alpha-helical domain"
 Misc-difference 887
 /note= "unidentified amino acid"
 Misc-difference 890
 /note= "unidentified amino acid"
 Peptide 1058..1059
 /label= linker_peptide
 Domain 1060..1171
 /label= TGF-beta-1
 /note= "human mature TGF-beta-1"

CA2151547-A.
 11-DEC-1995.
 12-JUN-1995; 95CA-02151547.
 10-JUN-1994; 94US-00259263.
 (USSU) US SURGICAL CORP.
 Gruskin EA, Espino P;
 WPI; 1996-140144/15.
 N-PSDB; AAT16516.

Chimaeric DNA encoding protein contg. extracellular matrix protein domain
 - and cellular regulatory factor domain, partic. useful as osteogenic

PT agents, also related vectors, transformed cells and chimaeric proteins.
 XX Disclosure; Fig 6; 59pp; English.
 CC A fusion protein (AAR89470) comprises the alpha-helical region of human
 CC collagen I(a) linked to the human mature transforming growth factor beta-
 CC 1 (TGF-beta-1). It can be expressed in Escherichia coli transformants
 CC carrying a vector incorporating a chimeric gene (AAT16516) coding for the
 CC fusion. The TGF-beta- moiety increases efficacy of the body's normal soft
 CC tissue repair response and also induces osteogenesis. The collagen moiety
 CC provides an integral substructure or scaffolding for the TGF and cells
 CC involved in reconstruction and growth. The fusion protein provides
 CC sustained release and delivery of TGF-beta-1 to a target tissue
 XX Sequence 1171 AA;
 SQ

Query Match 96.6%; Score 2966; DB 2; Length 1171;
 Best Local Similarity 95.6%; Pred. No. 3.1e-169;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GSEGEVGRGPPGPPGAGAPGADPCADGEPGAKGADGAPGAGPPGAGPPGSPG 60
 Db 201 GSEGEVGRGPPGPPGAGAPGADPCADGEPGAKGADGAPGAGPPGAGPPGSPG 260
 QY 61 GPGGPPGKDSGPPGAPGSKGDTGAKGPPGVGPPGPPGAGGPPGAGGPPGPPG 120
 Db 261 GPGGPPGKDSGPPGAPGSKGDTGAKGPPGVGPPGPPGAGGPPGAGGPPGPPG 320
 QY 121 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 180
 Db 321 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 380
 QY 181 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
 Db 381 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 440
 QY 241 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
 Db 441 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 500
 QY 301 GVPDGLGAPGSGARGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 360
 Db 501 GVPDGLGAPGSGARGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 560
 QY 361 GAPLEGMPGPPGAGAGLPDPKDRDAGPKGADGSPGKGVRLTGPDPGAGAGDK 420
 Db 561 GAPLEGMPGPPGAGAGLPDPKDRDAGPKGADGSPGKGVRLTGPDPGAGAGDK 620
 QY 421 GESGSPGAGTGPAGAPGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 480
 Db 621 GESGSPGAGTGPAGAPGPPGPPGAGPACGEEGPPGAGSPGEGLPAGPPGAGKPGEE 680
 QY 481 GPAGPAGPPGPPGAGAGKAGSAGPPGATGPPGAGRVGPPGPPGAGPPGPPGPA 540
 Db 681 GPAGPAGPPGPPGAGAGKAGSAGPPGATGPPGAGRVGPPGPPGAGPPGPPGPA 740
 QY 541 GKGE 544
 Db 741 GKGE 744

RESULT 11
 AAY84538
 ID AAY84538 standard; protein; 1171 AA.
 AC AAY84538;
 DT 25-JUL-2000 (first entry)
 DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;


```

AAR89471
ID AAR89471 standard; protein; 1388 AA.
XX
AC AAR89471;
XX
DT 01-OCT-1996 (first entry)
XX
DE Collagen/decorin fusion protein.
XX
KW Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
KW bone formation; tissue repair; fusion protein.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Domain 1..1057
FT /label= Collagen-IA
FT /note= "collagen IA alpha-helical domain"
FT Misc-difference 887
FT /note= "unidentified amino acid"
FT Misc-difference 890
FT /note= "unidentified amino acid"
FT Peptide 1058..1059
FT /label= Linker_peptide
FT Domain 1060..1388
FT /label= Decorin
XX
PN CA2151547-A.
XX
XX 11-DEC-1995.
XX
XX 12-JUN-1995; 95CA-02151547.
XX
XX 10-JUN-1994; 94US-00259263.
XX
XX (USSU ) US SURGICAL CORP.
XX
XX Gruskin EA, Espino P;
XX WPI; 1996-140144/15.
XX N-PSDB; AAT16517.
XX
XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
FT - and cellular regulatory factor domain, partic. useful as osteogenic
FT agents, also related vectors, transformed cells and chimaeric proteins.
XX
XX Disclosure; Fig 7; 59pp; English.
XX
XX A fusion protein (AAR89471) comprises the alpha-helical region of human
XX collagen I(a) linked to human dermatan sulphate proteoglycan (decorin).
XX It can be expressed in Escherichia coli transformants carrying a vector
XX incorporating a chimeric gene (AAT16517) coding for the fusion. The
XX decorin binds to type I collagen and thus affects Elbril formation. It
XX inhibits the cell attachment-promoting activity of collagen and
XX fibrinogen by binding to such molecules near their cell binding sites.
XX The collagen moiety provides an integral substratum or scaffolding for
XX the decorin. The fusion protein acts to reduce scarring of healing tissue
XX
SQ Sequence 1388 AA;

Query Match          96.6%; Score 2966; DB 2; Length 1388;
Best Local Similarity 95.6%; Pred. No. 3.6e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 CSEGGVGRGPPGPPAGAGPAGDPCADGEPGAKGADGAPGTAGAPGPPGARGPSQPE 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 201 CSEGGVGRGPPGPPGPPAGAGPAGNPGADGQPCAKGANGAPGTAGAPGPPGARGPSQ 260
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 GPGGPPGKGSDEPGSPKSGDGTAKGPGPVGVEGPPGAGEGKPGARGECPTGLP 120
DB 261 GPGGPPGKGSDEPGSPKSGDGTAKGPGPVGVEGPPGAGEGKPGARGECPTGLP 320
QY 121 GPPGRRGGGSRGPPGADGVAGPKPGAGRGSPGAGPKGSPGAGRPGEAGLPKAGLT 180

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 321 GPPGRRGGGSRGPPGADGVAGPKPGAGRGSPGAGPKGSPGAGRPGEAGLPKAGLT 380
QY 181 GSPGSPGPDGKTGTPPPAGEDGPPGPPGARGGAGVMGPPGKGAAGBPGKAGRGVP 240
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 381 GSPGSPGPDGKTGTPPPAGQDGRFPGPPGARGGAGVMGPPGKGAAGBPGKAGRGVP 440
QY 241 GPPGAVGAPGKDGAGAGEGPPGPPGAPGAGRGGERGPPGAGSPGEGLPGPAGPPGEGAKPGEE 300
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 441 GPPGAVGAPGKDGAGAGEGPPGPPGAPGAGRGGERGPPGAGSPGEGLPGPAGPPGEGAKPGEE 500
QY 301 GVPGLGAPGSPGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 501 GVPGLGAPGSPGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 560
QY 361 GAPGLEGMPPGARGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGTPTGPPGPPGAGPDK 420
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 561 GAPGLQMPGARGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGTPTGPPGPPGAGPDK 620
QY 421 GESGSPGAPGTGARGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 621 GESGSPGAPGTGARGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 680
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 681 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 740
QY 541 GKEG 544
   |||||
Db 741 GKEG 744

```

```

RESULT 15
AAY56800
ID AAY56800 standard; protein; 1411 AA.
XX
AC AAY56800;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human preproalpha 1 (I) collagen.
XX
KW Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin;
KW sequence selection and alignment domain; prosthetic implant; foodstuff;
KW medicine; type I collagen; human.
XX
OS Homo sapiens.
XX
PN EP967226-A2.
XX
PD 29-DEC-1999.
XX
PF 04-MAY-1999; 99EP-00303470.
XX
PR 08-MAY-1998; 98US-0084828P.
XX
PR 10-APR-1999; 99US-00289578.
XX
PA (COHE-) COHESION TECHNOLOGIES INC.
XX
PI Olsen DR, Hitzeman RA, Chisholm GE;
XX
DR WPI; 2000-074666/07.
XX
PT New method for production of fibrillar collagen, useful for preparing
PT telopeptide collagen fibrils and gelatin.
XX
XX Example 1; Fig 3A-B; 30pp; English.
XX
CC The invention provides a method for the production of fibrillar collagen.
CC The method comprises: (a) culturing a recombinant host cell comprising a
CC DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD
CC (sequence selection and alignment domain); and (b) producing the
CC fibrillar collagen. The methods are used to produce fibrillar collagen,

```


Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2966	96.6	1461	4	US-09-585-887-9		Sequence 9, Appli
2	2966	96.6	1461	4	US-09-289-578-9		Sequence 9, Appli
3	2966	96.6	1464	4	US-09-331-347C-21		Sequence 21, Appl
4	2857	93.1	822	3	US-09-219-849-49		Sequence 49, Appl
5	2842	92.6	1057	3	US-08-931-820-1		Sequence 1, Appli
6	2791.5	90.9	1341	3	US-08-963-825-18		Sequence 18, Appl
7	2791.5	90.9	1341	3	US-09-500-811-18		Sequence 18, Appl
8	2791.5	90.9	1341	3	US-09-570-573-18		Sequence 18, Appl
9	2791.5	90.9	1341	3	US-09-548-608-18		Sequence 18, Appl
10	2226	72.5	1017	4	US-08-468-996-10		Sequence 10, Appl
11	2202	71.7	1060	3	US-08-931-820-3		Sequence 3, Appli
12	2202	71.7	1418	3	US-08-963-825-20		Sequence 20, Appl
13	2202	71.7	1418	3	US-09-010-999-1		Sequence 1, Appli
14	2202	71.7	1418	3	US-09-500-811-20		Sequence 20, Appl
15	2202	71.7	1418	3	US-09-570-573-20		Sequence 20, Appl
16	2202	71.7	1418	3	US-09-548-608-20		Sequence 20, Appl
17	2192	71.4	1442	2	US-08-316-650-12		Sequence 12, Appl
18	2192	71.4	1442	5	PCT-US95-02251-12		Sequence 12, Appl
19	2156	70.2	595	3	US-09-219-849-48		Sequence 48, Appl
20	2156	70.2	595	3	US-09-219-849-50		Sequence 50, Appl
21	1901	61.9	1057	3	US-08-931-820-4		Sequence 4, Appli
22	1886	61.4	1024	3	US-08-931-820-2		Sequence 2, Appli
23	1886	61.4	1366	3	US-08-963-825-19		Sequence 19, Appl
24	1886	61.4	1366	3	US-09-500-811-19		Sequence 19, Appl
25	1886	61.4	1366	3	US-09-570-573-19		Sequence 19, Appl
26	1886	61.4	1366	3	US-09-548-608-19		Sequence 19, Appl
27	1885	61.4	1366	4	US-09-585-887-10		Sequence 10, Appl

QY 421 GESGSPAGTARGAPDRGEPGPPGAGPAGPAGDCEGAGKBERGDAGAKGDAGPP 480
Db 782 GESGSPAGTARGAPDRGEPGPPGAGPAGPAGDCEGAGKBERGDAGAKGDAGPP 841
QY 481 GPAGPAGPPGPIGVGAPGAPGAGSAGPPGATGPPGAGAGRVGPPGSGDAGPPGPPGPA 540
Db 842 GPAGPAGPPGPIGVGAPGAPGAGSAGPPGATGPPGAGAGRVGPPGSGDAGPPGPPGPA 901
QY 541 GKEG 544
Db 902 GKEG 905
RESULT 4
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUNSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49
Query Match 93.1%; Score 2857; DB 3; Length 822;
Best Local Similarity 91.4%; Pred. No. 4.6e-170;
Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;
QY 1 GSEGPVGRGPPGPPGAPGADPCGADGEPGAKGADGAPGAGPFGGARGPSGPE 60
Db 183 GSEGPVGRGPPGPPGAPGADPCGADGEPGAKGADGAPGAGPFGGARGPSGPE 242
QY 61 GPGPPGPKGDSGEPGAPGSKGDTCAKGPVGVGPPGAPGEGKPGARGEPGPTGLP 120
Db 243 GPGPPGPKGDSGEPGAPGSKGDTCAKGPVGVGPPGAPGEGKPGARGEPGPTGLP 302
QY 121 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 180
Db 303 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 362
QY 181 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 240
Db 363 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 422
QY 241 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 300
Db 423 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 482
QY 301 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 360
Db 483 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 542
QY 361 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 420

Db 543 GAPGLQMPGERGAAGLPGGKDRGDAGPKGADGSPKGDGARGLTGPTGPPGAPAGPDK 602
QY 421 GESGSPAGTARGAPDRGEPGPPGAGPAGPAGDCEGAGKBERGDAGAKGDAGPP 480
Db 603 GEAGSPGPPGPTGARGAPDRGEPGPPGAGPAGPAGDCEGAGKBERGDAGAKGDAGPP 662
QY 481 GPAGPAGPPGPIGVGAPGAPGAGSAGPPGATGPPGAGAGRVGPPGSGDAGPPGPPGPA 540
Db 663 GPAGPAGPPGPIGVGAPGAPGAGSAGPPGATGPPGAGAGRVGPPGSGDAGPPGPPGPA 722
QY 541 GKEG 544
Db 723 GKEG 726
RESULT 5
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1
Query Match 92.6%; Score 2842; DB 3; Length 1057;
Best Local Similarity 91.9%; Pred. No. 4.9e-169;
Matches 500; Conservative 24; Mismatches 20; Indels 0; Gaps 0;
QY 1 GSEGPVGRGPPGPPGAPGADPCGADGEPGAKGADGAPGAGPFGGARGPSGPE 60
Db 201 GSEGPVGRGPPGPPGAPGADPCGADGEPGAKGADGAPGAGPFGGARGPSGPE 260
QY 61 GPGPPGPKGDSGEPGAPGSKGDTCAKGPVGVGPPGAPGEGKPGARGEPGPTGLP 120
Db 261 GPGPPGPKGDSGEPGAPGSKGDTCAKGPVGVGPPGAPGEGKPGARGEPGPTGLP 320
QY 121 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 180
Db 321 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 380
QY 181 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 240
Db 381 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 440
QY 241 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 300
Db 441 GPPGRRGPPGRRGPPGADGAVGKPGPAGERSGPPGAPGKSGPGEAGRPCEAGLPCAAGLT 500


```

; APPLICATION NUMBER: 08/187,319
; FILING DATE: 08/187,319
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-527-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-500-811-18

```

```

Query Match          90.9%; Score 2791.5; DB 3; Length 1341;
Best Local Similarity 87.6%; Pred. No. 8.2e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

QY 1 GSEGEVGRGPPGPPGAGAGPAGDPGADGE---PGAKGADGAPGIAGAPGPGAR--- 54
DB 211 GASGPMGPRGPPGPPGKBGBZAGKPCRPGZRGPPGPGZGARGLPCTAGLPGMKHGRGFS 270
QY 55 -----GPSGPEG-----PG--GPRPKGDSGEPGAPGSKGDTGAKGEPGV 93
DB 271 GLBGAKGABGAPGAGKZGPGSCZGAGPMGPPGPKGNSGEPGAPGSKGDTGAKGEPGV 330
QY 94 GVEGPPGAPGEGKPGARGEPTGLPGLPPGERGPGSGRPGPGADGVAGPKGPAGERGSP 153
DB 331 GVQGPAPGAGEGKPGARGEPTGLPGLPPGERGPGSGRPGPGADGVAGPKGPAGERGSP 390
QY 154 GPAGPKGSPGABRGEAGLPGAKGLTGSPPSGPDGKTPGPPGAGEDRGPFGPPGPGAR 213
DB 391 GPAGPKGSPGABRGEAGLPGAKGLTGSPPSGPDGKTPGPPGAGEDRGPFGPPGPGAR 450
QY 214 GEAGVMPGPKGAGEPKGAGERGVPPGPGAVGPKAGKDGAGAGPPGPPGAPAGERGEE 273
DB 451 GQAGVMGPPGPKGAGEPKGAGERGVPPGPGAVGPKAGKDGAGAGQGGPPGPPGAPAGERGEQ 510
QY 274 GPAGSPGEGLPGPAGPGEAGKPGCEGVPGDLGAPGPGSARGEPGPFGERGVGPGPA 333
DB 511 GPAGSPGEGLPGPAGPGEAGKPGCEGVPGDLGAPGPGSARGEPGPFGERGVGPGPA 570
QY 334 GPPGADGAPGDDGAKGDAGAPGAPGSEGAPGLEGMPPGERGAAGLPGPKGDRGDAGPKGAD 393
DB 571 GPRGANGAPGNDGAKGDAGAPGAPGSCQAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGAD 630
QY 394 GSPGKDGVRGLTGPTGPPGPGAPGPDGKESGSPGAPGTGARGAPGDRGERGPPGPGAPGA 453
DB 631 GSPGKDGVRGLTGPTGPPGPGAPGPDGKESGSPGAPGTGARGAPGDRGERGPPGPGAPGA 690
QY 454 GPPGADGPPGAKGEPGDAGKGDAGPPGAPGAPGPGTGDVGAPGAKGARGSPGPGAT 513
DB 691 GPPGADGPPGAKGEPGDAGKGDAGPPGAPGAPGPGTGDVGAPGAKGARGSPGPGAT 750
QY 514 GPPGAGAGRVGPPGSGDAGPPGPPGAPGKEG 544
DB 751 GPPGAGAGRVGPPGSGNAGPPGPPGAPGKEG 781

```

RESULT 8

```

US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per

```

```

; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-570-573-18

```

```

Query Match          90.9%; Score 2791.5; DB 3; Length 1341;
Best Local Similarity 87.6%; Pred. No. 8.2e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

QY 1 GSEGEVGRGPPGPPGAGAGPAGDPGADGE---PGAKGADGAPGIAGAPGPGAR--- 54
DB 211 GASGPMGPRGPPGPPGKBGBZAGKPCRPGZRGPPGPGZGARGLPCTAGLPGMKHGRGFS 270
QY 55 -----GPSGPEG-----PG--GPRPKGDSGEPGAPGSKGDTGAKGEPGV 93
DB 271 GLBGAKGABGAPGAGKZGPGSCZGAGPMGPPGPKGNSGEPGAPGSKGDTGAKGEPGV 330
QY 94 GVEGPPGAPGEGKPGARGEPTGLPGLPPGERGPGSGRPGPGADGVAGPKGPAGERGSP 153
DB 331 GVQGPAPGAGEGKPGARGEPTGLPGLPPGERGPGSGRPGPGADGVAGPKGPAGERGSP 390
QY 154 GPAGPKGSPGABRGEAGLPGAKGLTGSPPSGPDGKTPGPPGAGEDRGPFGPPGPGAR 213
DB 391 GPAGPKGSPGABRGEAGLPGAKGLTGSPPSGPDGKTPGPPGAGEDRGPFGPPGPGAR 450
QY 214 GEAGVMPGPKGAGEPKGAGERGVPPGPGAVGPKAGKDGAGAGPPGPPGAPAGERGEE 273
DB 451 GQAGVMGPPGPKGAGEPKGAGERGVPPGPGAVGPKAGKDGAGAGQGGPPGPPGAPAGERGEQ 510
QY 274 GPAGSPGEGLPGPAGPGEAGKPGCEGVPGDLGAPGPGSARGEPGPFGERGVGPGPA 333
DB 511 GPAGSPGEGLPGPAGPGEAGKPGCEGVPGDLGAPGPGSARGEPGPFGERGVGPGPA 570

```


; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-468-996-10

Query Match 72.5%; Score 2226; DB 4; Length 1017;
Best Local Similarity 72.6%; Pred. No. 7.6e-131;
Matches 395; Conservative 39; Mismatches 110; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPAGAGPAGDPCADGDEPGKAGDAGPAGAGPFGPCARGPSPE 60
DB 184 GPEGAGQGRGPPGPPAGAGPAGDPCADGDEPGKAGDAGPAGAGPFGPCARGPSPE 243
QY 61 GPGGPPGKDSGSPGAPGSGKDTCAKGEPPGVGPPGPPAGBEGKPGARGEPGPTGLP 120
DB 244 GATGLPFGKGTGKPGIAGFKGEQPKGEPGPPAGPQAGPAGBEGKPGARGEPGVP 303
QY 121 GPPGRRGPPGRRGPPGADGVAGPKGPPAGERSPPGAPGKSPGSEAGRPGEAGLP 180
DB 304 GPPGRRGPPGRRGPPGADGVAGPKGPPAGERSPPGAPGKSPGSEAGRPGEAGLP 363
QY 181 GSPGSPGPPGKGTGPPGPPAGDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
DB 364 GRPDGAGPQGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 423
QY 241 GPPGAVGPPGKDGAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
DB 424 GAPLGLGPPGKDGAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 483
QY 301 GVPGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 484 GVPGEAGPGLVPRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 543
QY 361 GAPLGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
DB 544 GPPGLQGMPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 603
QY 421 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
DB 604 GEVGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 663
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
DB 664 GPQPSGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 723
QY 541 GKEG 544
DB 724 GKDG 727

RESULT 11
US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT: Assay for collagen degradation
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931, 820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1

; FILING DATE:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
US-08-931-820-3

Query Match 71.7%; Score 2202; DB 3; Length 1060;
Best Local Similarity 71.9%; Pred. No. 2.4e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPAGAGPAGDPCADGDEPGKAGDAGPAGAGPFGPCARGPSPE 60
DB 203 GPEGAGQGRGPPGPPAGAGPAGDPCADGDEPGKAGDAGPAGAGPFGPCARGPSPE 262
QY 61 GPGGPPGKDSGSPGAPGSGKDTCAKGEPPGVGPPGPPAGBEGKPGARGEPGPTGLP 120
DB 263 GATGLPFGKGTGKPGIAGFKGEQPKGEPGPPAGPQAGPAGBEGKPGARGEPGVP 322
QY 121 GPPGRRGPPGRRGPPGADGVAGPKGPPAGERSPPGAPGKSPGSEAGRPGEAGLP 180
DB 323 GPPGRRGPPGRRGPPGADGVAGPKGPPAGERSPPGAPGKSPGSEAGRPGEAGLP 382
QY 181 GSPGSPGPPGKGTGPPGPPAGDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
DB 383 GRPDGAGPQGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 442
QY 241 GPPGAVGPPGKDGAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
DB 443 GAPLGLGPPGKDGAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 502
QY 301 GVPGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 503 GVPGEAGPGLVPRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 562
QY 361 GAPLGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
DB 563 GPPGLQGMPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 622
QY 421 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
DB 623 GEVGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 682
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
DB 683 GPQPSGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 742
QY 541 GKEG 544
DB 743 GKDG 746

RESULT 12
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC

;
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
; US-09-570-573-20

Query Match 71.7%; Score 2202; DB 3; Length 1418;
Best local Similarity 71.9%; Pred. No. 3.1e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

QY 1 GSEGEVVRGEPGPPGAGAGDAGDGEFGAKGADGAGIAGAPGPPGARGSPGEE 60
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 315 GPEGAAQPRGEPGTPGSPGAGASGNGTGTGIPGAKGSAGAGIAGAPGPPGPPDPQ 374
QY 61 GPGGPPGPKGDSGEPGAPGSGKGTGAKGEPGVGSPGPPGAGEGKPGARGEPGPTGLP 120
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 375 GATGPLGPKQTKPKGTAGKQGKGPPGAPGQAFGAPGEGRGARGEPGPGVGT 434
QY 121 GPPGERGPGSGRFPAGDGVAGPKGAPGERSPGAPGKSGFCEAGRPGAGLPGAKGLT 180
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 435 GPPGERGAPGNRGFPQDGLAGPKGAPGERGPGSLAGPKGANGDPGRPGEPGLPGARGLT 494
QY 181 GSPGSPGPDGKTGPPGAGEDGPPGPPGARGGAGVMPGPKGAAGEPKAGRGVVP 240
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 495 GRFDAGPQKVGPSGAPGEDGPPGPPGQARGQPGVNGFPFGKANGEPGKAGEKGLP 554
QY 241 GPPGAVGPKDGEAGAGPPGPPGAPGRRGEGPAGSPGFGLPGPAGPPGCEAGKPGEE 300
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 555 GAPLRLGLPKDGETGAEGPPGPPGAPGERGQAGFPSPGQGLPGPPFPGEKPKPDQ 614
QY 301 GVPDGLGAPGSPGARGEPGPPGPPGPPGADGAPGDDGAKGDAGAPGAPGSE 360
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 615 GVFEAGAFGLVGRGERGFPGERGSPGAGLQGLQGRGLPGTGTGPKGASGAPGPPGAQ 674
QY 361 GAPLLEMPGERGAAGLPGPKDGRDAGPKGADGSPGKDVGLTGTPTGPPGAPAGDK 420
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 675 GPPGLOGMFGERGAGIAGPKDGRDVGKGPGEAGPDKDGGRLGTPTGPPGAPANGEK 734
QY 421 GSGSPGPPAGPTGARGAPGDRGEPGPPGAPGAGDGEPPGAKGEPGDAGAKGDAGPP 480
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 735 GEVGGPPGAGSAGARGAPGERGTGPPGTSGTAGPPGADGQFGAKGEQGEAGQKGDAGAP 794
QY 481 GPAGPAGPPGPTGVDYCAPKAGKARGSAGPPGATGTFPGAAGRVGPPGSGDAGPPGPPA 540
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 795 GPQPSGAFPGQGTGVTGPKARGAQSPPGATGFPGAAGRVGPPGSGNFGPPGPPGFS 854

QY 541 GKEG 544
Db 855 GKDG 858

Search completed: June 17, 2005, 15:17:02
Job time : 25.0908 secs

Query Match	100.0%;	Score 3070;	DB 17;	Length 544;
Best Local Similarity	100.0%;	Pred. NO. 4e-144;		
Matches 544;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0
Qy	1	GSFGPGEVRGEP	PPGAGAGDGDGEC	PKAGKADGAGGTAGAPFPFGARGPSGPE 60
Db	1	GSFGPGEVRGEP	PPGAGAGDGDGEC	PKAGKADGAGGTAGAPFPFGARGPSGPE 60
Qy	61	GPGGPFPKDKSDG	PGAPGSGKGTGAKG	EPFVGVEPPGAGBEGKPGARGEPGPTGLP 120

QY 241 GPPCAVGPAGKDGAGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 300
DB 424 GPPCAVGPAGKDGAGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 483
QY 301 GVPDILGAPGSGARGEPGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 360
DB 484 GVPDILGAPGSGARGEPGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 543
QY 361 GAPLEGMPGSRGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 420
DB 544 GAPLEGMPGSRGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 603
QY 421 GESGSPGAGTARGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 480
DB 604 GESGSPGAGTARGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 663
QY 481 GPAGPAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 540
DB 664 GPAGPAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 723
QY 541 GKEG 544
DB 724 GKEG 727

RESULT 4

US-10-104-889-16
; Sequence 16, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

Query Match 96.6%; Score 2966; DB 15; Length 1057;
Best Local Similarity 95.6%; Pred. No. 8.9e-139;

Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPGAGPAGDPCAGCEPGKAGDAGPAGPAGPPGAGPPGPPG 60
DB 201 GSEGEVGRGPPGPPGAGPAGDPCAGCEPGKAGDAGPAGPAGPPGAGPPGPPG 260
QY 61 GPGGPPGKDSGPPGAGPAGSGKGTGAGKGPVGVGPPGPPGAGPAGPPGPPG 120
DB 261 GPGGPPGKDSGPPGAGPAGSGKGTGAGKGPVGVGPPGPPGAGPAGPPGPPG 320
QY 121 GPPGERGPPGPPGAGPAGSGKGTGAGKGPVGVGPPGPPGAGPAGPPGPPG 180
DB 321 GPPGERGPPGPPGAGPAGSGKGTGAGKGPVGVGPPGPPGAGPAGPPGPPG 380
QY 181 GSPGPPGKDSGPPGAGPAGSGKGTGAGKGPVGVGPPGPPGAGPAGPPGPPG 240
DB 381 GSPGPPGKDSGPPGAGPAGSGKGTGAGKGPVGVGPPGPPGAGPAGPPGPPG 440
QY 241 GPPCAVGPAGKDGAGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 300
DB 441 GPPCAVGPAGKDGAGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 500
QY 301 GVPDILGAPGSGARGEPGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 360
DB 501 GVPDILGAPGSGARGEPGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 560
QY 361 GAPLEGMPGSRGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 420
DB 561 GAPLEGMPGSRGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 620
QY 421 GESGSPGAGTARGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 480
DB 621 GESGSPGAGTARGAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 680
QY 481 GPAGPAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 540
DB 681 GPAGPAGPPGPPGAGPAGCERGGEPAGSPGEGGLPGPAGPPGEGAGKPGEE 740
QY 541 GKEG 544
DB 741 GKEG 744

RESULT 5

US-10-104-889-20
; Sequence 20, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768

QY 541 GKEG 544
 Db 741 GKEG 744

RESULT 7
 US-10-104-889-6
 ; Sequence 6, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GRUSKIN, ELLIOT A.
 ; BROKAW, JANE
 ; ZHANG, GUANGHUI
 ; PAOLELLA, DAVID
 ; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DILWORTH & BARRESE
 ; STREET: 333 EARLE OVINGTON BOULEVARD
 ; CITY: UNIONDALE
 ; STATE: NY
 ; COUNTRY: U.S.A.
 ; ZIP: 11553
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/104,889
 ; FILING DATE: 22-Mar-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/169,768
 ; FILING DATE: 09-OCT-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEEN, JEFFREY S
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 228-8484
 ; TELEFAX: (516) 228-8516
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1169 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

RESULT 8
 US-10-104-889-8
 ; Sequence 8, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GRUSKIN, ELLIOT A.
 ; BROKAW, JANE
 ; ZHANG, GUANGHUI
 ; PAOLELLA, DAVID
 ; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DILWORTH & BARRESE
 ; STREET: 333 EARLE OVINGTON BOULEVARD
 ; CITY: UNIONDALE
 ; STATE: NY
 ; COUNTRY: U.S.A.
 ; ZIP: 11553
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/104,889
 ; FILING DATE: 22-Mar-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/169,768
 ; FILING DATE: 09-OCT-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEEN, JEFFREY S
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 228-8484
 ; TELEFAX: (516) 228-8516
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1171 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-10-104-889-8
 Query Match 96.6%; Score 2966; DB 15; Length 1171;
 Best Local Similarity 95.6%; Pred. No. 9.6e-139;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

Query Match 96.6%; Score 2966; DB 15; Length 1169;
 Best Local Similarity 95.6%; Pred. No. 9.6e-139;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPGAGAGPAGDPCADCEPGKAGKADGAPGIAGAPGPPGARGPSGPE 60
 Db 201 GSEGPQVGRGPPGPPGAGAGPAGNPGADGQPGKANGAPGIAGAPGPPGARGPSGPQ 260

QY 61 GPGGPPGKGSBPGAPGSGKDTGAKGEPGVGVGPPGPPGAGBEGKPGARGEPGPTGLP 120
 Db 261 GPGGPPGKGSBPGAPGSGKDTGAKGEPGVGVGPPGPPGAGBEGKPGARGEPGPTGLP 320

QY 121 GPPGCGGSGRGGPAGDVGAPKGPAGERSGPPAGKPSGPGAGRPGAGLPCAKGLT 180
 Db 321 GPPGCGGSGRGGPAGDVGAPKGPAGERSGPPAGKPSGPGAGRPGAGLPCAKGLT 380

QY 181 GSPGSGPDGKTGPPGPPGAGEGRGPPGPPGARGAGVGMGPPGKGAAGEPKAGRGVP 240
 Db 381 GSPGSGPDGKTGPPGPPGAGEGRGPPGPPGARGAGVGMGPPGKGAAGEPKAGRGVP 440

QY 241 GPPGAVGPAGKDGAEAGBPPGPPGAPGAGERGEGPAGSPGFGLFPGAPPPGAEAKPGEE 300

Qy	1	GSEFGEVGRSEPGPPGPPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGPPFGARGPSGPE	60
Db	201	GSEGPQVGRSEPGPPGPPAGAAGPAGNFGADGQPGAKGANGAPGIAGAPGPPFGARGPSGPQ	260
Qy	61	GPQGPQPKDSEPGAPSGKDTGAKGBPGVPVGEPPGPAGBEGKPGARGEPGPTGLP	120
Db	261	GPQGPQPKNSGEPGAPSGKDTGAKGBPGVPVQVQGPGPAGBEGKPGARGEPGPTGLP	320
Qy	121	GPQERGGPSRGFPGADGVAGPKGPAGBERGSPGAPGKSPGACGPPGACGILPGAKGLT	180
Db	321	GPQERGGPSRGFPGADGVAGPKGPAGBERGSPGAPGKSPGACGPPGACGILPGAKGLT	380
Qy	181	GSPGSPGDKTGPFPAGEDGRPGPPPPARGBAGVMGPPGKGAAGSPGKAGRGVP	240
Db	381	GSPGSPGDKTGPFPAGASQDGRPGPPPPARGAQAGVMGPPGKGAAGSPGKAGRGVP	440
Qy	241	GPPGAVVPAGKDEAGAEOPPPGAPGABERGEBSGPFEGILPGPAGPPGBAGKPGSE	300
Db	441	GPPGAVVPAGKDEAGAEOPPPGAPGABERGEBSGPFEGILPGPAGPPGBAGKPGSQ	500
Qy	301	GVPPDLGAPGPGSARGEPGPFGERGVGPPGPPGADGNAPGDGAKGDAGAPGAPGSE	360
Db	501	GVTPEDLGAPGPGSARGERGFPGERGVQPPGPPGANGAPGNDGAKGDAGAPGAPGSG	560
Qy	361	GAPGLEMPGERGAAGLPKPKDGRDGDAPKGDGSPGKGVRLTGPFGPPGAPGAPGDK	420
Db	561	GAPGLQMPGERGAAGLPKPKDGRDGDAPKGDGSPGKGVRLTGPFGPPGAPGAPGDK	620
Qy	421	GESRPSGPAGPTGARGAPCDRGEPPGPPGAPGFPAGDGEPCAKGEPGDAGKGDGPP	480
Db	621	GESRPSGPAGPTGARGAPCDRGEPPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPG	680
Qy	481	GPAGAPGPPGPTGDVGAPOAKGARGSGAPGCATGFPCCAGRVGPPGSDGAPGPPGPPA	540
Db	681	GPAGAPGPPGPTLONVGAPOAKGARGSGAPGCATGFPCCAGRVGPPGSDGAPGPPGPPA	740
Qy	541	GKXE 544	
Db	741	GKEX 744	

RESULT 9

US-10-104-889-10
; Sequence 10, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998


```
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-468-091-25

Query Match      96.6%; Score 2966; DB 16; Length 1461;
Best Local Similarity 95.6%; Pred. No. 1.1e-138;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGEPGPPGAGAGPAGDPCADGEPGAKGADGAPGIAGAPGFGARGPSGPE 60
    |||||
DB 359 GSEGEVGRGEPGPPGAGAGPAGNPGADGQPGAKGANGAPGIAGAPGFGARGPSGQ 418
    |||||

QY 61 GPGGPPGKDSGSGPAGPSKGDTCAGKEPGVGVGPPGAGERGKPGARGEPGPTGLP 120
    |||||
DB 419 GPGGPPGKDSGSGPAGPSKGDTCAGKEPGVGVGPPGAGERGKPGARGEPGPTGLP 478
    |||||

QY 121 GPPGRGPGSGRGGFFPCADGVAGPKPAGERSGPGAGPKGSPGEAGRPGEAGLPCAKGLT 180
    |||||
DB 479 GPPGRGPGSGRGGFFPCADGVAGPKPAGERSGPGAGPKGSPGEAGRPGEAGLPCAKGLT 538
    |||||

QY 181 GSPGSPGDKTGTGPPGAGEDGPPGPPGARGGAGVWGPFGKGAAGEPGKAGERGVP 240
    |||||
DB 539 GSPGSPGDKTGTGPPGAGEDGPPGPPGARGGAGVWGPFGKGAAGEPGKAGERGVP 598
    |||||

QY 241 GPPGAVGPAKDGEGAGAGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 300
    |||||
DB 599 GPPGAVGPAKDGEGAGAGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 658
    |||||

QY 301 GVPDLGAPGSGARGEPGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 360
    |||||
DB 659 GVPDLGAPGSGARGEPGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 718
    |||||

QY 361 GAPLEGWPGERGAAGLPGPKDGDGAPKGDGSGKDGVRGLTGTGPPGAGAPGDK 420
    |||||
DB 719 GAPLEGWPGERGAAGLPGPKDGDGAPKGDGSGKDGVRGLTGTGPPGAGAPGDK 778
    |||||

QY 421 GESGSPGAGTGTARGAPGDRGEPGPPGAGPAGPPGAGPPGAGPPGAGPPGAGPP 480
    |||||
DB 779 GESGSPGAGTGTARGAPGDRGEPGPPGAGPAGPPGAGPPGAGPPGAGPPGAGPP 838
    |||||

QY 481 GPAGPAGPPGPIGVDVGAAGKAGSAGPPGATGPPGAAGRVGPPGSPGDAGPPGPPGA 540
    |||||
DB 839 GPAGPAGPPGPIGVDVGAAGKAGSAGPPGATGPPGAAGRVGPPGSPGDAGPPGPPGA 898
    |||||

QY 541 GKEG 544
    |||||
DB 899 GKEG 902
    |||||
```

RESULT 11

```
US-09-918-715-261
; Sequence 261, Application US/09918715
; Publication No. US2003001757A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 261
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-261

Query Match      96.6%; Score 2966; DB 10; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1.1e-138;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGEPGPPGAGAGPAGDPCADGEPGAKGADGAPGIAGAPGFGARGPSGPE 60
    |||||
DB 362 GSEGEVGRGEPGPPGAGAGPAGNPGADGQPGAKGANGAPGIAGAPGFGARGPSGQ 421
    |||||

QY 61 GPGGPPGKDSGSGPAGPSKGDTCAGKEPGVGVGPPGAGERGKPGARGEPGPTGLP 120
    |||||
DB 422 GPGGPPGKDSGSGPAGPSKGDTCAGKEPGVGVGPPGAGERGKPGARGEPGPTGLP 481
    |||||

QY 121 GPPGRGPGSGRGGFFPCADGVAGPKPAGERSGPGAGPKGSPGEAGRPGEAGLPCAKGLT 180
    |||||
DB 482 GPPGRGPGSGRGGFFPCADGVAGPKPAGERSGPGAGPKGSPGEAGRPGEAGLPCAKGLT 541
    |||||

QY 181 GSPGSPGDKTGTGPPGAGEDGPPGPPGARGGAGVWGPFGKGAAGEPGKAGERGVP 240
    |||||
DB 542 GSPGSPGDKTGTGPPGAGEDGPPGPPGARGGAGVWGPFGKGAAGEPGKAGERGVP 601
    |||||

QY 241 GPPGAVGPAKDGEGAGAGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 300
    |||||
DB 602 GPPGAVGPAKDGEGAGAGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 661
    |||||

QY 301 GVPDLGAPGSGARGEPGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 360
    |||||
DB 662 GVPDLGAPGSGARGEPGPPGPPGAGPAGERGEEGPPGAGPPGAGPPGAGPPGEE 721
    |||||

QY 361 GAPLEGWPGERGAAGLPGPKDGDGAPKGDGSGKDGVRGLTGTGPPGAGAPGDK 420
    |||||
DB 722 GAPLEGWPGERGAAGLPGPKDGDGAPKGDGSGKDGVRGLTGTGPPGAGAPGDK 781
    |||||

QY 421 GESGSPGAGTGTARGAPGDRGEPGPPGAGPAGPPGAGPPGAGPPGAGPPGAGPP 480
    |||||
DB 782 GESGSPGAGTGTARGAPGDRGEPGPPGAGPAGPPGAGPPGAGPPGAGPPGAGPP 841
    |||||

QY 481 GPAGPAGPPGPIGVDVGAAGKAGSAGPPGATGPPGAAGRVGPPGSPGDAGPPGPPGA 540
    |||||
DB 842 GPAGPAGPPGPIGVDVGAAGKAGSAGPPGATGPPGAAGRVGPPGSPGDAGPPGPPGA 901
    |||||

QY 541 GKEG 544
    |||||
DB 902 GKEG 905
    |||||
```

RESULT 12

```
US-10-060-036-159
; Sequence 159, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Daxin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
```


US-10-060-036-159

```

Query Match          96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1.1e-138;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPGAGAGAGDPCAGDGEPCAGKAGDAGPAGIAGAPGPPGARGSPGE 60
DB 362 GSEGPQVGRGPPGPPGAGAGAGDPCAGDGEPCAGKAGDAGPAGIAGAPGPPGARGSPGE 421
QY 61 GPGGPPGKDSBEPGAPGSKGDTGAKGEPGVPVGEPPGAGGEGKAGRGEPGPTGLP 120
DB 422 GPGGPPGKNSBEPGAPGSKGDTGAKGEPGVPVGEPPGAGGEGKAGRGEPGPTGLP 481
QY 121 GPPGERGPPGPPGAGDGVAGPKGAGERGSPGAPGKSPGGEAGRPGEAGLPGAKGLT 180
DB 482 GPPGERGPPGPPGAGDGVAGPKGAGERGSPGAPGKSPGGEAGRPGEAGLPGAKGLT 541
QY 181 GSPGSPGDKTGTGPPGAGDGRPPGPPGARGEGAGVGMFPKGAAGEPPKAGRGVVP 240
DB 542 GSPGSPGDKTGTGPPGAGDGRPPGPPGARGEGAGVGMFPKGAAGEPPKAGRGVVP 601
QY 241 GPPGAVGPAGKDEAGAGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 300
DB 602 GPPGAVGPAGKDEAGAGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 661
QY 301 GVPDGLGAPGSGARGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 360
DB 662 GVPDGLGAPGSGARGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 721
QY 361 GAPLEGMPGERGAGLPGPKGDRGDAGPKGADGSGKDGVRGLTGTPIGPPGAPGDK 420
DB 722 GAPLEGMPGERGAGLPGPKGDRGDAGPKGADGSGKDGVRGLTGTPIGPPGAPGDK 781
QY 421 GESGPPGAGTARGAGPPGPPGAGPPGAGPPGAGDGEPCAGKGERGPDAGKGDAGPP 480
DB 782 GESGPPGAGTARGAGPPGPPGAGPPGAGPPGAGDGEPCAGKGERGPDAGKGDAGPP 841
QY 481 GPAGPAGPPGPIGVNAGPAGKARGAGSAGPPGATGFPGAAGRVGPPGPPGPPGPPGPA 540
DB 842 GPAGPAGPPGPIGVNAGPAGKARGAGSAGPPGATGFPGAAGRVGPPGPPGPPGPPGPA 901
QY 541 GKEG 544
DB 902 GKEG 905

```

RESULT 13

```

US-10-171-311-36
; Sequence 36, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamakaz, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238

```

; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 36
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-36

```

Query Match 96.6%; Score 2966; DB 14; Length 1464;

```

Best Local Similarity 95.6%; Pred. No. 1.1e-138;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 GSEGEVGRGPPGPPGAGAGAGDPCAGDGEPCAGKAGDAGPAGIAGAPGPPGARGSPGE 60
DB 362 GSEGPQVGRGPPGPPGAGAGAGDPCAGDGEPCAGKAGDAGPAGIAGAPGPPGARGSPGE 421
QY 61 GPGGPPGKDSBEPGAPGSKGDTGAKGEPGVPVGEPPGAGGEGKAGRGEPGPTGLP 120
DB 422 GPGGPPGKNSBEPGAPGSKGDTGAKGEPGVPVGEPPGAGGEGKAGRGEPGPTGLP 481
QY 121 GPPGERGPPGPPGAGDGVAGPKGAGERGSPGAPGKSPGGEAGRPGEAGLPGAKGLT 180
DB 482 GPPGERGPPGPPGAGDGVAGPKGAGERGSPGAPGKSPGGEAGRPGEAGLPGAKGLT 541
QY 181 GSPGSPGDKTGTGPPGAGDGRPPGPPGARGEGAGVGMFPKGAAGEPPKAGRGVVP 240
DB 542 GSPGSPGDKTGTGPPGAGDGRPPGPPGARGEGAGVGMFPKGAAGEPPKAGRGVVP 601
QY 241 GPPGAVGPAGKDEAGAGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 300
DB 602 GPPGAVGPAGKDEAGAGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 661
QY 301 GVPDGLGAPGSGARGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 360
DB 662 GVPDGLGAPGSGARGPPGPPGAGRGGERGPPGAGSGPFEGLPKGAAGEPPKAGRGVVP 721
QY 361 GAPLEGMPGERGAGLPGPKGDRGDAGPKGADGSGKDGVRGLTGTPIGPPGAPGDK 420
DB 722 GAPLEGMPGERGAGLPGPKGDRGDAGPKGADGSGKDGVRGLTGTPIGPPGAPGDK 781
QY 421 GESGPPGAGTARGAGPPGPPGAGPPGAGPPGAGDGEPCAGKGERGPDAGKGDAGPP 480
DB 782 GESGPPGAGTARGAGPPGPPGAGPPGAGPPGAGDGEPCAGKGERGPDAGKGDAGPP 841
QY 481 GPAGPAGPPGPIGVNAGPAGKARGAGSAGPPGATGFPGAAGRVGPPGPPGPPGPPGPA 540
DB 842 GPAGPAGPPGPIGVNAGPAGKARGAGSAGPPGATGFPGAAGRVGPPGPPGPPGPPGPA 901
QY 541 GKEG 544
DB 902 GKEG 905

```

RESULT 14

```

US-10-216-705-21
; Sequence 21, Application US/10216705
; Publication No. US20030096973A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met
; FILE OF INVENTION: Obtaining Such and Their Uses
; FILE REFERENCE: 1149-3 DIV
; CURRENT APPLICATION NUMBER: US/10/216,705
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 09/331,347
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-216-705-21

```


Query Match	96.6%	Score 2966	DB 14	Length 1464
Best Local Similarity	95.6%	Pred. No. 1.1e-138		
Matches 520	Conservative 21	Mismatches 3	Indels 0	Gaps 0
Qy	1	GSEGEVGRGEPGPPGAGAAAGADPCADGCEGPAKAGADGAPGTAGAPFCFCAAGPSGPE	60	
Db	362	GSEGPQGVGEFPGPPGAGAAAGPAGNPGADQPCAKGANGAPGTAGAGFFFCARGPSGPQ	421	
Qy	61	GPQPPGPKGDSGSPGAPGSGKDTGAKGEPGVGVEGPPGAGBEGKPGARGEPPTGLP	120	
Db	422	GPQPPGPKGNSGSEFGAPGSGKDTGAKGEPGVGQGPFGAGEGKRGARGEPPTGLP	481	
Qy	121	GPPCERGPGSRGPPGADGVAGPKGPACERGSPPGAPGKSPGEAGRPGEAGLPCAKGLT	180	
Db	482	GPPCERGPGSRGPPGADGVAGPKGPACERGSPPGAPGKSPGEAGRPGEAGLPCAKGLT	541	
Qy	181	GSPGSPGDKTGGPPGAGEDRGFPGPPGARGGAGVMGPPGPKGAAGEPKAGERGVP	240	
Db	542	GSPGSPGDKTGGPPGAGEDRGFPGPPGARGGAGVMGPPGPKGAAGEPKAGERGVP	601	
Qy	241	GPPGAVGPAKDXGAGAGGPPGPPGAPGAGERGEEGPPAGSPGPEGLPQAPAGPPGCAKPGEE	300	
Db	602	GPPGAVGPAKDXGAGAGGPPGPPGAPGAGERGEEGPPAGSPGPEGLPQAPAGPPGCAKPGEE	661	
Qy	301	GVPDGLGAPGSGARGEPGFPCEGVEGPPGAPGPPGADGAPDDGAKGDAGCAPGSE	360	
Db	662	GVPDGLGAPGSGARGEPGFPCEGVEGPPGAPGPPGADGAPDDGAKGDAGCAPGSE	721	
Qy	361	GAPGLEGMGPERGAAGLPGPKGDRGDAGPKGADGSPGKDXGVRGLTGPTGPPGAPAGDGK	420	
Db	722	GAPGLQGMGPERGAAGLPGPKGDRGDAGPKGADGSPGKDXGVRGLTGPTGPPGAPAGDGK	781	
Qy	421	GESGSPGAGPTGARGAPCDRGEPGPPGAPGPPGADGSPGAKGEPDAGAKGDAGPP	480	
Db	782	GESGSPGAGPTGARGAPCDRGEPGPPGAPGPPGADGSPGAKGEPDAGAKGDAGPP	841	
Qy	481	GPAGPAGPPGPTGIVGAPGKAGARGAGPPGATGPPGAAGRGVPPGPPGSDAGPPGPPGPA	540	
Db	842	GPAGPAGPPGPTGIVGAPGKAGARGAGPPGATGPPGAAGRGVPPGPPGSDAGPPGPPGPA	901	
Qy	541	GKEG 544		
Db	902	GKEG 905		

```

RESULT 15
US-10-149-352-2
; Sequence 2, Application US/10149352
; Publication No. US20030105050A1
; GENERAL INFORMATION:
; APPLICANT: Berl, Rajinder
; TITLE OF INVENTION: ANTIGENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/10/149,352
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-149-352-2

Query Match          96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1,1e-138;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps
0

Qy      1  GSEGPVGRBPGPPGAGAGAGDPGADGPGAGKADGAPGACGAGPGFPGARGPSGPE 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: June 17, 2005, 15:35:49
Job time : 61.5892 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:42:55 ; Search time 18.6216 seconds
(without alignments)
2810.812 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGFGEVGRGEPGPAGG.....PPSGDAGPPGPGACGEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: Pirl:*
2: Pirl:*
3: Pirl:*
4: Pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1464	1 CGHUL5	collagen alpha 1(I)
2	2857	93.1	1453	2 S21626	collagen alpha 1(I)
3	2775	90.4	1042	1 CGCH15	collagen alpha 1(I)
4	2220	72.3	1418	2 T45467	collagen alpha 1(I)
5	2202	71.7	1487	1 CGHU6C	collagen alpha 1(I)
6	2189	71.3	1419	2 A41182	collagen alpha 1(I)
7	2189	71.3	1487	2 B41182	collagen alpha 1(I)
8	2154	70.2	671	1 CGRT1S	collagen alpha 1(I)
9	2147	69.9	1486	1 B40333	collagen alpha 1(I)
10	2131	69.4	1492	2 A40333	collagen alpha 1(I)
11	2076.5	67.6	779	1 CGH01S	collagen alpha 1(I)
12	1921.5	62.6	1496	1 CGH02V	collagen alpha 2(I)
13	1915	62.4	1373	1 A43291	collagen alpha 2(I)
14	1910	62.2	1497	2 I49607	procollagen type V
15	1908	62.1	1464	2 S59856	collagen alpha 1(I)
16	1905.5	62.1	1049	1 CGH07S	collagen alpha 1(I)
17	1901	61.9	1466	1 CGH07L	collagen alpha 1(I)
18	1886	61.4	1366	2 I50694	collagen alpha 2(I)
19	1881	61.3	886	2 I50694	collagen alpha 2(I)
20	1550.5	50.5	1414	1 S23809	collagen alpha 1(I)
21	1457.5	47.5	1838	1 CGHU1V	collagen alpha 1(I)
22	1444	47.0	1806	1 CGHU1V	collagen alpha 1(I)
23	1432.5	46.7	1843	2 S18803	collagen alpha 1(I)
24	1418.5	46.2	1027	2 S28774	collagen alpha 1(I)
25	1413	46.0	964	1 CGCH2S	collagen alpha 2(I)
26	1411.5	46.0	1024	2 S18251	collagen alpha 1(I)
27	1400	45.6	888	2 S28791	collagen alpha 1(I)
28	1399.5	45.6	1546	1 CGHU2E	collagen alpha 2(I)
29	1376	44.8	1691	1 S22917	collagen alpha 5(I)

30 1375.5 44.8 2944 2 A54849 collagen alpha 1(I)
31 1372 44.7 1763 2 S16366 collagen alpha 2(I)
32 1364.5 44.4 1549 2 T48103 type VII collagen
33 1357.5 44.2 1690 1 CGHU1B collagen alpha 4(I)
34 1339.5 43.6 1603 2 S23810 collagen alpha 1(I)
35 1339.5 43.6 1670 1 CGHU3B collagen alpha 3(I)
36 1335.5 43.5 920 2 A45748 collagen alpha 1(I)
37 1320 43.0 1669 1 CGHU4B collagen alpha 1(I)
38 1310.5 42.7 1758 2 T29350 hypothetical prote
39 1310 42.7 1669 1 CGMS4B collagen alpha 1(I)
40 1309 42.6 1759 2 T29351 collagen alpha 2(I)
41 1303.5 42.5 812 2 S31521 collagen COLF1 - f
42 1286.5 41.9 1712 1 CGHU2B collagen alpha 2(I)
43 1279.5 41.7 754 2 A55267 collagen alpha 5(I)
44 1277 41.6 921 2 S42617 collagen alpha 1(I)
45 1276.5 41.6 1142 2 JX0369 collagen alpha 1(X)

ALIGNMENTS

RESULT 1

CGHU1S
collagen alpha 1(I) chain precursor - human
N:Alternate names: procollagen alpha 1(I) chain
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 09-Jul-2004
C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1135269; A29439; I53466; A02852; I37247
R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988
A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex
A:Reference number: I60114; MUID:88329734; PMID:2843432
A:Accession: I60114
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369, 'L', 371-589 <DAL>
A:Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP
R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock
Biochem. J. 253, 919-922, 1988
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t
A:Reference number: S01143; MUID:89025644; PMID:3178743
A:Accession: S01143
A:Molecule type: mRNA
A:Residues: 1-472 <TRO>
A:Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NID:
A>Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.
Nature 310, 337-340, 1984
A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A:Reference number: A93335; MUID:84270697; PMID:6462220
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58, 'Q', 60-181 <CHU>
A:Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA25394.1; PID:G35658
R:Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.
J. Biol. Chem. 262, 15151-15157, 1987
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh
A:Reference number: I55254; MUID:88033098; PMID:2822714
A:Accession: I55254
A>Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control of
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
 A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter structure and organization of the 5' and 3' ends only are shown; mutant sequence 263-Arg reported
 A;Reference number: I55237; MUID:85130970; PMID:2857713
 A;Accession: I55237
 A;Status: translation not shown; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-34 <CH2>
 A;Cross-references: GB:M10627; NID:g180383; PIDN:AAA51992.1; PID:g553226
 R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steifmann, B.; Rao, V.H.; Hollister, J. Biol. Chem. 265, 6312-6317, 1990
 A;Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-terminal propeptides
 A;Reference number: A35233; MUID:90202908; PMID:2318855
 A;Accession: A35233
 A;Molecule type: protein
 A;Residues: 33-52 <WIR>
 A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved propeptide
 R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F. EMBO J. 8, 1705-1710, 1989
 A;Title: A base substitution in the exon of a collagen gene causes alternative splicing
 A;Reference number: S09400; MUID:89356643; PMID:2767050
 A;Accession: S09400
 A;Molecule type: mRNA
 A;Residues: 156-183 <WEI>
 R;Click, E.M.; Bornstein, P. Biochemistry 9, 4699-4706, 1970
 A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) procollagen chain
 A;Reference number: A90567; MUID:71038625; PMID:5529814
 A;Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
 A;Accession: B90567
 A;Molecule type: protein
 A;Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z', 235, 'Z', 238, 'Z', 240-241, 'Z', 243-244, 'Z', 246-247, 'Z', 249-250, 'Z', 252-253, 'Z', 255-256, 'Z', 258-259, 'Z', 261-262, 'Z', 264-265, 'Z', 267-268, 'Z', 270-271, 'Z', 273-274, 'Z', 276-277, 'Z', 279-280, 'Z', 282-283, 'Z', 285-286, 'Z', 288-289, 'Z', 291-292, 'Z', 294-295, 'Z', 297-298, 'Z', 300-301, 'Z', 303-304, 'Z', 306-307, 'Z', 309-310, 'Z', 312-313, 'Z', 315-316, 'Z', 318-319, 'Z', 321-322, 'Z', 324-325, 'Z', 327-328, 'Z', 330-331, 'Z', 333-334, 'Z', 336-337, 'Z', 339-340, 'Z', 342-343, 'Z', 345-346, 'Z', 348-349, 'Z', 351-352, 'Z', 354-355, 'Z', 357-358, 'Z', 360-361, 'Z', 363-364, 'Z', 366-367, 'Z', 369-370, 'Z', 372-373, 'Z', 375-376, 'Z', 378-379, 'Z', 381-382, 'Z', 384-385, 'Z', 387-388, 'Z', 390-391, 'Z', 393-394, 'Z', 396-397, 'Z', 399-400, 'Z', 402-403, 'Z', 405-406, 'Z', 408-409, 'Z', 411-412, 'Z', 414-415, 'Z', 417-418, 'Z', 420-421, 'Z', 423-424, 'Z', 426-427, 'Z', 429-430, 'Z', 432-433, 'Z', 435-436, 'Z', 438-439, 'Z', 441-442, 'Z', 444-445, 'Z', 447-448, 'Z', 450-451, 'Z', 453-454, 'Z', 456-457, 'Z', 459-460, 'Z', 462-463, 'Z', 465-466, 'Z', 468-469, 'Z', 471-472, 'Z', 474-475, 'Z', 477-478, 'Z', 480-481, 'Z', 483-484, 'Z', 486-487, 'Z', 489-490, 'Z', 492-493, 'Z', 495-496, 'Z', 498-499, 'Z', 501-502, 'Z', 504-505, 'Z', 507-508, 'Z', 510-511, 'Z', 513-514, 'Z', 516-517, 'Z', 519-520, 'Z', 522-523, 'Z', 525-526, 'Z', 528-529, 'Z', 531-532, 'Z', 534-535, 'Z', 537-538, 'Z', 540-541, 'Z', 543-544, 'Z', 546-547, 'Z', 549-550, 'Z', 552-553, 'Z', 555-556, 'Z', 558-559, 'Z', 561-562, 'Z', 564-565, 'Z', 567-568, 'Z', 570-571, 'Z', 573-574, 'Z', 576-577, 'Z', 579-580, 'Z', 582-583, 'Z', 585-586, 'Z', 588-589, 'Z', 591-592, 'Z', 594-595, 'Z', 597-598, 'Z', 600-601, 'Z', 603-604, 'Z', 606-607, 'Z', 609-610, 'Z', 612-613, 'Z', 615-616, 'Z', 618-619, 'Z', 621-622, 'Z', 624-625, 'Z', 627-628, 'Z', 630-631, 'Z', 633-634, 'Z', 636-637, 'Z', 639-640, 'Z', 642-643, 'Z', 645-646, 'Z', 648-649, 'Z', 651-652, 'Z', 654-655, 'Z', 657-658, 'Z', 660-661, 'Z', 663-664, 'Z', 666-667, 'Z', 669-670, 'Z', 672-673, 'Z', 675-676, 'Z', 678-679, 'Z', 681-682, 'Z', 684-685, 'Z', 687-688, 'Z', 690-691, 'Z', 693-694, 'Z', 696-697, 'Z', 699-700, 'Z', 702-703, 'Z', 705-706, 'Z', 708-709, 'Z', 711-712, 'Z', 714-715, 'Z', 717-718, 'Z', 720-721, 'Z', 723-724, 'Z', 726-727, 'Z', 729-730, 'Z', 732-733, 'Z', 735-736, 'Z', 738-739, 'Z', 741-742, 'Z', 744-745, 'Z', 747-748, 'Z', 750-751, 'Z', 753-754, 'Z', 756-757, 'Z', 759-760, 'Z', 762-763, 'Z', 765-766, 'Z', 768-769, 'Z', 771-772, 'Z', 774-775, 'Z', 777-778, 'Z', 780-781, 'Z', 783-784, 'Z', 786-787, 'Z', 789-790, 'Z', 792-793, 'Z', 795-796, 'Z', 798-799, 'Z', 801-802, 'Z', 804-805, 'Z', 807-808, 'Z', 810-811, 'Z', 813-814, 'Z', 816-817, 'Z', 819-820, 'Z', 822-823, 'Z', 825-826, 'Z', 828-829, 'Z', 831-832, 'Z', 834-835, 'Z', 837-838, 'Z', 840-841, 'Z', 843-844, 'Z', 846-847, 'Z', 849-850, 'Z', 852-853, 'Z', 855-856, 'Z', 858-859, 'Z', 861-862, 'Z', 864-865, 'Z', 867-868, 'Z', 870-871, 'Z', 873-874, 'Z', 876-877, 'Z', 879-880, 'Z', 882-883, 'Z', 885-886, 'Z', 888-889, 'Z', 891-892, 'Z', 894-895, 'Z', 897-898, 'Z', 900-901, 'Z', 903-904, 'Z', 906-907, 'Z', 909-910, 'Z', 912-913, 'Z', 915-916, 'Z', 918-919, 'Z', 921-922, 'Z', 924-925, 'Z', 927-928, 'Z', 930-931, 'Z', 933-934, 'Z', 936-937, 'Z', 939-940, 'Z', 942-943, 'Z', 945-946, 'Z', 948-949, 'Z', 951-952, 'Z', 954-955, 'Z', 957-958, 'Z', 960-961, 'Z', 963-964, 'Z', 966-967, 'Z', 969-970, 'Z', 972-973, 'Z', 975-976, 'Z', 978-979, 'Z', 981-982, 'Z', 984-985, 'Z', 987-988, 'Z', 990-991, 'Z', 993-994, 'Z', 996-997, 'Z', 999-1000, 'Z', 1002-1003, 'Z', 1005-1006, 'Z', 1008-1009, 'Z', 1011-1012, 'Z', 1014-1015, 'Z', 1017-1018, 'Z', 1020-1021, 'Z', 1023-1024, 'Z', 1026-1027, 'Z', 1029-1030, 'Z', 1032-1033, 'Z', 1035-1036, 'Z', 1038-1039, 'Z', 1041-1042, 'Z', 1044-1045, 'Z', 1047-1048, 'Z', 1050-1051, 'Z', 1053-1054, 'Z', 1056-1057, 'Z', 1059-1060, 'Z', 1062-1063, 'Z', 1065-1066, 'Z', 1068-1069, 'Z', 1071-1072, 'Z', 1074-1075, 'Z', 1077-1078, 'Z', 1080-1081, 'Z', 1083-1084, 'Z', 1086-1087, 'Z', 1089-1090, 'Z', 1092-1093, 'Z', 1095-1096, 'Z', 1098-1099, 'Z', 1101-1102, 'Z', 1104-1105, 'Z', 1107-1108, 'Z', 1110-1111, 'Z', 1113-1114, 'Z', 1116-1117, 'Z', 1119-1120, 'Z', 1122-1123, 'Z', 1125-1126, 'Z', 1128-1129, 'Z', 1131-1132, 'Z', 1134-1135, 'Z', 1137-1138, 'Z', 1140-1141, 'Z', 1143-1144, 'Z', 1146-1147, 'Z', 1149-1150, 'Z', 1152-1153, 'Z', 1155-1156, 'Z', 1158-1159, 'Z', 1161-1162, 'Z', 1164-1165, 'Z', 1167-1168, 'Z', 1170-1171, 'Z', 1173-1174, 'Z', 1176-1177, 'Z', 1179-1180, 'Z', 1182-1183, 'Z', 1185-1186, 'Z', 1188-1189, 'Z', 1191-1192, 'Z', 1194-1195, 'Z', 1197-1198, 'Z', 1200-1201, 'Z', 1203-1204, 'Z', 1206-1207, 'Z', 1209-1210, 'Z', 1212-1213, 'Z', 1215-1216, 'Z', 1218-1219, 'Z', 1221-1222, 'Z', 1224-1225, 'Z', 1227-1228, 'Z', 1230-1231, 'Z', 1233-1234, 'Z', 1236-1237, 'Z', 1239-1240, 'Z', 1242-1243, 'Z', 1245-1246, 'Z', 1248-1249, 'Z', 1251-1252, 'Z', 1254-1255, 'Z', 1257-1258, 'Z', 1260-1261, 'Z', 1263-1264, 'Z', 1266-1267, 'Z', 1269-1270, 'Z', 1272-1273, 'Z', 1275-1276, 'Z', 1278-1279, 'Z', 1281-1282, 'Z', 1284-1285, 'Z', 1287-1288, 'Z', 1290-1291, 'Z', 1293-1294, 'Z', 1296-1297, 'Z', 1299-1300, 'Z', 1302-1303, 'Z', 1305-1306, 'Z', 1308-1309, 'Z', 1311-1312, 'Z', 1314-1315, 'Z', 1317-1318, 'Z', 1320-1321, 'Z', 1323-1324, 'Z', 1326-1327, 'Z', 1329-1330, 'Z', 1332-1333, 'Z', 1335-1336, 'Z', 1338-1339, 'Z', 1341-1342, 'Z', 1344-1345, 'Z', 1347-1348, 'Z', 1350-1351, 'Z', 1353-1354, 'Z', 1356-1357, 'Z', 1359-1360, 'Z', 1362-1363, 'Z', 1365-1366, 'Z', 1368-1369, 'Z', 1371-1372, 'Z', 1374-1375, 'Z', 1377-1378, 'Z', 1380-1381, 'Z', 1383-1384, 'Z', 1386-1387, 'Z', 1389-1390, 'Z', 1392-1393, 'Z', 1395-1396, 'Z', 1398-1399, 'Z', 1401-1402, 'Z', 1404-1405, 'Z', 1407-1408, 'Z', 1410-1411, 'Z', 1413-1414, 'Z', 1416-1417, 'Z', 1419-1420, 'Z', 1422-1423, 'Z', 1425-1426, 'Z', 1428-1429, 'Z', 1431-1432, 'Z', 1434-1435, 'Z', 1437-1438, 'Z', 1440-1441, 'Z', 1443-1444, 'Z', 1446-1447, 'Z', 1449-1450, 'Z', 1452-1453, 'Z', 1455-1456, 'Z', 1458-1459, 'Z', 1461-1462, 'Z', 1464-1465, 'Z', 1467-1468, 'Z', 1470-1471, 'Z', 1473-1474, 'Z', 1476-1477, 'Z', 1479-1480, 'Z', 1482-1483, 'Z', 1485-1486, 'Z', 1488-1489, 'Z', 1491-1492, 'Z', 1494-1495, 'Z', 1497-1498, 'Z', 1500-1501, 'Z', 1503-1504, 'Z', 1506-1507, 'Z', 1509-1510, 'Z', 1512-1513, 'Z', 1515-1516, 'Z', 1518-1519, 'Z', 1521-1522, 'Z', 1524-1525, 'Z', 1527-1528, 'Z', 1530-1531, 'Z', 1533-1534, 'Z', 1536-1537, 'Z', 1539-1540, 'Z', 1542-1543, 'Z', 1545-1546, 'Z', 1548-1549, 'Z', 1551-1552, 'Z', 1554-1555, 'Z', 1557-1558, 'Z', 1560-1561, 'Z', 1563-1564, 'Z', 1566-1567, 'Z', 1569-1570, 'Z', 1572-1573, 'Z', 1575-1576, 'Z', 1578-1579, 'Z', 1581-1582, 'Z', 1584-1585, 'Z', 1587-1588, 'Z', 1590-1591, 'Z', 1593-1594, 'Z', 1596-1597, 'Z', 1599-1600, 'Z', 1602-1603, 'Z', 1605-1606, 'Z', 1608-1609, 'Z', 1611-1612, 'Z', 1614-1615, 'Z', 1617-1618, 'Z', 1620-1621, 'Z', 1623-1624, 'Z', 1626-1627, 'Z', 1629-1630, 'Z', 1632-1633, 'Z', 1635-1636, 'Z', 1638-1639, 'Z', 1641-1642, 'Z', 1644-1645, 'Z', 1647-1648, 'Z', 1650-1651, 'Z', 1653-1654, 'Z', 1656-1657, 'Z', 1659-1660, 'Z', 1662-1663, 'Z', 1665-1666, 'Z', 1668-1669, 'Z', 1671-1672, 'Z', 1674-1675, 'Z', 1677-1678, 'Z', 1680-1681, 'Z', 1683-1684, 'Z', 1686-1687, 'Z', 1689-1690, 'Z', 1692-1693, 'Z', 1695-1696, 'Z', 1698-1699, 'Z', 1701-1702, 'Z', 1704-1705, 'Z', 1707-1708, 'Z', 1710-1711, 'Z', 1713-1714, 'Z', 1716-1717, 'Z', 1719-1720, 'Z', 1722-1723, 'Z', 1725-1726, 'Z', 1728-1729, 'Z', 1731-1732, 'Z', 1734-1735, 'Z', 1737-1738, 'Z', 1740-1741, 'Z', 1743-1744, 'Z', 1746-1747, 'Z', 1749-1750, 'Z', 1752-1753, 'Z', 1755-1756, 'Z', 1758-1759, 'Z', 1761-1762, 'Z', 1764-1765, 'Z', 1767-1768, 'Z', 1770-1771, 'Z', 1773-1774, 'Z', 1776-1777, 'Z', 1779-1780, 'Z', 1782-1783, 'Z', 1785-1786, 'Z', 1788-1789, 'Z', 1791-1792, 'Z', 1794-1795, 'Z', 1797-1798, 'Z', 1800-1801, 'Z', 1803-1804, 'Z', 1806-1807, 'Z', 1809-1810, 'Z', 1812-1813, 'Z', 1815-1816, 'Z', 1818-1819, 'Z', 1821-1822, 'Z', 1824-1825, 'Z', 1827-1828, 'Z', 1830-1831, 'Z', 1833-1834, 'Z', 1836-1837, 'Z', 1839-1840, 'Z', 1842-1843, 'Z', 1845-1846, 'Z', 1848-1849, 'Z', 1851-1852, 'Z', 1854-1855, 'Z', 1857-1858, 'Z', 1860-1861, 'Z', 1863-1864, 'Z', 1866-1867, 'Z', 1869-1870, 'Z', 1872-1873, 'Z', 1875-1876, 'Z', 1878-1879, 'Z', 1881-1882, 'Z', 1884-1885, 'Z', 1887-1888, 'Z', 1890-1891, 'Z', 1893-1894, 'Z', 1896-1897, 'Z', 1899-1900, 'Z', 1902-1903, 'Z', 1905-1906, 'Z', 1908-1909, 'Z', 1911-1912, 'Z', 1914-1915, 'Z', 1917-1918, 'Z', 1920-1921, 'Z', 1923-1924, 'Z', 1926-1927, 'Z', 1929-1930, 'Z', 1932-1933, 'Z', 1935-1936, 'Z', 1938-1939, 'Z', 1941-1942, 'Z', 1944-1945, 'Z', 1947-1948, 'Z', 1950-1951, 'Z', 1953-1954, 'Z', 1956-1957, 'Z', 1959-1960, 'Z', 1962-1963, 'Z', 1965-1966, 'Z', 1968-1969, 'Z', 1971-1972, 'Z', 1974-1975, 'Z', 1977-1978, 'Z', 1980-1981, 'Z', 1983-1984, 'Z', 1986-1987, 'Z', 1989-1990, 'Z', 1992-1993, 'Z', 1995-1996, 'Z', 1998-1999, 'Z', 2001-2002, 'Z', 2004-2005, 'Z', 2007-2008, 'Z', 2010-2011, 'Z', 2013-2014, 'Z', 2016-2017, 'Z', 2019-2020, 'Z', 2022-2023, 'Z', 2025-2026, 'Z', 2028-2029, 'Z', 2031-2032, 'Z', 2034-2035, 'Z', 2037-2038, 'Z', 2040-2041, 'Z', 2043-2044, 'Z', 2046-2047, 'Z', 2049-2050, 'Z', 2052-2053, 'Z', 2055-2056, 'Z', 2058-2059, 'Z', 2061-2062, 'Z', 2064-2065, 'Z', 2067-2068, 'Z', 2070-2071, 'Z', 2073-2074, 'Z', 2076-2077, 'Z', 2079-2080, 'Z', 2082-2083, 'Z', 2085-2086, 'Z', 2088-2089, 'Z', 2091-2092, 'Z', 2094-2095, 'Z', 2097-2098, 'Z', 2100-2101, 'Z', 2103-2104, 'Z', 2106-2107, 'Z', 2109-2110, 'Z', 2112-2113, 'Z', 2115-2116, 'Z', 2118-2119, 'Z', 2121-2122, 'Z', 2124-2125, 'Z', 2127-2128, 'Z', 2130-2131, 'Z', 2133-2134, 'Z', 2136-2137, 'Z', 2139-2140, 'Z', 2142-2143, 'Z', 2145-2146, 'Z', 2148-2149, 'Z', 2151-2152, 'Z', 2154-2155, 'Z', 2157-2158, 'Z', 2160-2161, 'Z', 2163-2164, 'Z', 2166-2167, 'Z', 2169-2170, 'Z', 2172-2173, 'Z', 2175-2176, 'Z', 2178-2179, 'Z', 2181-2182, 'Z', 2184-2185, 'Z', 2187-2188, 'Z', 2190-2191, 'Z', 2193-2194, 'Z', 2196-2197, 'Z', 2199-2200, 'Z', 2202-2203, 'Z', 2205-2206, 'Z', 2208-2209, 'Z', 2211-2212, 'Z', 2214-2215, 'Z', 2217-2218, 'Z', 2220-2221, 'Z', 2223-2224, 'Z', 2226-2227, 'Z', 2229-2230, 'Z', 2232-2233, 'Z', 2235-2236, 'Z', 2238-2239, 'Z', 2241-2242, 'Z', 2244-2245, 'Z', 2247-2248, 'Z', 2250-2251, 'Z', 2253-2254, 'Z', 2256-2257, 'Z', 2259-2260, 'Z', 2262-2263, 'Z', 2265-2266, 'Z', 2268-2269, 'Z', 2271-2272, 'Z', 2274-2275, 'Z', 2277-2278, 'Z', 2280-2281, 'Z', 2283-2284, 'Z', 2286-2287, 'Z', 2289-2290, 'Z', 2292-2293, 'Z', 2295-2296, 'Z', 2298-2299, 'Z', 2301-2302, 'Z', 2304-2305, 'Z', 2307-2308, 'Z', 2310-2311, 'Z', 2313-2314, 'Z', 2316-2317, 'Z', 2319-2320, 'Z', 2322-2323, 'Z', 2325-2326, 'Z', 2328-2329, 'Z', 2331-2332, 'Z', 2334-2335, 'Z', 2337-2338, 'Z', 2340-2341, 'Z', 2343-2344, 'Z', 2346-2347, 'Z', 2349-2350, 'Z', 2352-2353, 'Z', 2355-2356, 'Z', 2358-2359, 'Z', 2361-2362, 'Z', 2364-2365, 'Z', 2367-2368, 'Z', 2370-2371, 'Z', 2373-2374, 'Z', 2376-2377, 'Z', 2379-2380, 'Z', 2382-2383, 'Z', 2385-2386, 'Z', 2388-2389, 'Z', 2391-2392, 'Z', 2394-2395, 'Z', 2397-2398, 'Z', 2400-2401, 'Z', 2403-2404, 'Z', 2406-2407, 'Z', 2409-2410, 'Z', 2412-2413, 'Z', 2415-2416, 'Z', 2418-2419, 'Z', 2421-2422, 'Z', 2424-2425, 'Z', 2427-2428, 'Z', 2430-2431, 'Z', 2433-2434, 'Z', 2436-2437, 'Z', 2439-2440, 'Z', 2442-2443, 'Z', 2445-2446, 'Z', 2448-2449, 'Z', 2451-2452, 'Z', 2454-2455, 'Z', 2457-2458, 'Z', 2460-2461, 'Z', 2463-2464, 'Z', 2466-2467, 'Z', 2469-2470, 'Z', 2472-2473, 'Z', 2475-2476, 'Z', 2478-2479, 'Z', 2481-2482, 'Z', 2484-2485, 'Z', 2487-2488, 'Z', 2490-2491, 'Z', 2493-2494, 'Z', 2496-2497, 'Z', 2499-2500, 'Z', 2502-2503, 'Z', 2505-2506, 'Z', 2508-2509, 'Z', 2511-2512, 'Z', 2514-2515, 'Z', 2517-2518, 'Z', 2520-2521, 'Z', 2523-2524, 'Z', 2526-2527, 'Z', 2529-2530, 'Z', 2532-2533, 'Z', 2535-2536, 'Z', 2538-2539, 'Z', 2541-2542, 'Z', 2544-2545, 'Z', 2547-2548, 'Z', 2550-2551, 'Z', 2553-2554, 'Z', 2556-2557, 'Z', 2559-2560, 'Z', 2562-2563, 'Z', 2565-2566, 'Z', 2568-2569, 'Z', 2571-2572, 'Z', 2574-2575, 'Z', 2577-2578, 'Z', 2580-2581, 'Z', 2583-2584, 'Z', 2586-2587, 'Z', 2589-2590, 'Z', 2592-2593, 'Z', 2595-2596, 'Z', 2598-2599, 'Z', 2601-2602, 'Z', 2604-2605, 'Z', 2607-2608, 'Z', 2610-2611, 'Z', 2613-2614, 'Z', 2616-2617, 'Z', 2619-2620, 'Z', 2622-2623, 'Z', 2625-2626, 'Z', 2628-2629, 'Z', 2631-2632, 'Z', 2634-2635, 'Z', 2637-2638, 'Z', 2640-2641, 'Z', 2643-2644, 'Z', 2646-2647, 'Z', 2649-2650, 'Z', 2652-2653, 'Z', 2655-2656, 'Z', 2658-2659, 'Z', 2661-2662, 'Z', 2664-2665, 'Z', 2667-2668, 'Z', 2670-2671, 'Z', 2673-2674, 'Z', 2676-2677, 'Z', 2679-2680, 'Z', 2682-2683, 'Z', 2685-2686, 'Z', 2688-2689, 'Z', 2691-2692, 'Z', 2694-2695, 'Z', 2697-2698, 'Z', 2700-2701, 'Z', 2703-2704, 'Z', 2706-2707, 'Z', 2709-2710, 'Z', 2712-2713, 'Z', 2715-2716, 'Z', 2718-2719, 'Z', 2721-2722, 'Z', 2724-2725, 'Z', 2727-2728, 'Z', 2730-2731, 'Z', 2733-2734, 'Z', 2736-2737, 'Z', 2739-2740, 'Z', 2742-2743, 'Z', 2745-2746, 'Z', 2748-2749, 'Z', 2751-2752, 'Z', 2754-2755, 'Z', 2757-2758, 'Z', 2760-2761, 'Z', 2763-2764, 'Z', 2766-2767, 'Z', 2769-2770, 'Z', 2772-2773, 'Z', 2775-2776, 'Z', 2778-277


```
QY 181 GSPGSPGDKTGPDPAGEDGRPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 240
Db 531 GSPGSPGDKTGPDPAGEDGRPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGLP 590
QY 241 GPPGAVGAGKDGAGAGGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 300
Db 591 GPPGAVGAGKDGAGAGGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 439
QY 301 GVPDLGAPGSGARGPGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 360
Db 651 GVPDLGAPGSGARGPGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 710
QY 361 GAGLEGMPGPPGARGPGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 420
Db 711 GAGLEGMPGPPGARGPGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 770
QY 421 GESGSPGAGTARGAGPPGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 480
Db 771 GESGSPGAGTARGAGPPGPPGPPGPPGARGVAGVMPGPKGAAGEPGKAGERGVP 830
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
Db 831 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 590
QY 541 GKEG 544
Db 891 GKEG 894

RESULT 3
CGCHIS
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C:Species: Gallus gallus (chicken)
C>Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
C:Accession: A90458; A90181; A02857
R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A:Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete primary
A:Reference number: A90458; PMID:82231995; PMID:7093229
A:Accession: A90458
A:Molecule type: protein
A:Residues: 1-1036 <HIG>
A:Experimental source: skin
A:Note: this is the latest in a series of papers from these workers elucidating the sequ
R:Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
A:Reference number: A90181; PMID:72243016; PMID:5047697
A:Accession: A90181
A:Molecule type: protein
A:Residues: 1037-1042 <BYR>
A:Experimental source: skin
A:Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
C:Comment: Most of the prolines at the third position of the tripeptide repeating unit
C:Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in p
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglycamic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 90.4%; Score 2775; DB 1; Length 1042;
Best Local Similarity 89.0%; Pred. No. 1.2e-126;
Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

QY 1 GSEGPVGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60
Db 200 GSEGPVGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 259
QY 61 GPGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 120
Db 260 GPGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 319
QY 121 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 180
```

```
Db 320 GPAGRGAGSGRGFPDGIAGPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 379
QY 181 GSPGSPGDKTGPDPAGEDGRPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240
Db 380 GSPGSPGDKTGPDPAGEDGRPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 439
QY 241 GPPGAVGAGKDGAGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
Db 440 GPPGAVGAGKDGAGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 499
QY 301 GVPDLGAPGSGARGPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
Db 500 GVPDLGAPGSGARGPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 559
QY 361 GAGLEGMPGPPGARGPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 420
Db 560 GAGLEGMPGPPGARGPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 619
QY 421 GESGSPGAGTARGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
Db 620 GESGSPGAGTARGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 679
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
Db 680 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 739
QY 541 GKEG 544
Db 740 GKZG 743

RESULT 4
T45467
collagen alpha 1(I) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C:Accession: T45467
R:Richardson, D.W.; Dodge, G.R.
Submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in eq
A:Reference number: 222977
A:Accession: T45467
A:Molecule type: mRNA
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-1418 <RIC>
A:Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:AA805773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 72.3%; Score 2220; DB 2; Length 1418;
Best Local Similarity 72.6%; Pred. No. 6.5e-100;
Matches 395; Conservative 37; Mismatches 112; Indels 0; Gaps 0;

QY 1 GSEGPVGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60
Db 315 GPEAGQGPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 374
QY 61 GPGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 120
Db 375 GATGFLGPKQQTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 434
QY 121 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 180
Db 435 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 494
QY 181 GSPGSPGDKTGPDPAGEDGRPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240
Db 495 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 554
QY 241 GPPGAVGAGKDGAGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
Db 555 GAGPLGRLPKDGTGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 614
```


A,Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A>Note: the GenBank FID is based on an incorrect reading frame
A:Accession: I37250
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 541-560 <SANS>
A,Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A:Accession: I37251

Query Match 71.7%; Score 2202; DB 1; Length 1487;
Best Local Similarity 71.9%; Pred. No. 4.9e-99;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

QY 1 GSEPGGVGRCPPGGPAGAGPADPDGADGEPKAKGADGAPGTAGAPFGFARGPSGPE 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 384 GPEAQOGRPEGPTPGSPGASGNGTDTGICPAKGSAGAPGTAGAPFGPRGPDPQ 443
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 GPQGPPPKDSDSEPAPSGSKDGTGAKGERPQVCGVEPPPACGEKPGKARGECPTGLP 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 444 GATGTLGPKQTGKTGTAGTGKGQGRGEPGAPGAPAGEEKKRGARGGPGVPI 503
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 GPPERGGSGRFPFGADGVAGVKPAGERGSPGAPKGPSGEAGRPFEGALPGA KGLT 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 504 GPRGERGAPGNRGFFPGQDGLAGPKGAPGERGSPSLAGPKANGDPGRGEPGLPGARGLT 563
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 GPSGSPDKCTGPPGAPGEDGRPGPPGPGARCEAGVMGFPGKGAAGRPKAGERVVP 240
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 564 GRFDAGPQKVQPSGAPGEDGRPGPPGPGQARGQPQGMVFPPGKANGEPFKAGEKGLP 623
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 GPFCAVGPAKGDGAEAGPPGAPGAGERGEGPAGSPGFEGLPGPAGPPGEAGKPGBE 300
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 624 GAPCLRLPKCKGETGAEGPPGAPGAGERGEQAGPCPSFGQLPGPPPGEGKEFGDQ 683
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 GVPGDLGACPSPARGEPPPERGVEGPPGAPPGADGAPGDGAKGDAGAPGAPSE 360
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 684 GVPEAGAPLVPRGERGPPGERGSPGAGLQGLQGRGLPGTPTDGPFGKASGAPGPPGAQ 743
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 GAPGLEMPGERAAGLPGPKDRGDAGPKGADGSPKDGVRGLTGPIGPPGPAGAPGDK 420
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 744 GPFGLQMPGERCAAGIAGPKDGRDVGEKGPSEA-PKDGGRGLTGPIGPPGPAGANGEK 803
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 421 GESGSGPPAGPTGARGAPDRGDRGPPGPPAGFAGPGADGEPGAKGEPGDAGAKGDAGPP 480
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 804 GEYGPPGPPAGSAGARGAPERGTGTPPTSGTAGPPCADGQPGAKGEQGEAGQKDAGAP 863
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 481 GPAGPAGPPGIIDVCAPAKGARGSAGPPGATGTFPCAAGRVGPPCPGSDAGPPGPPGA 540
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 864 GPQPPSSAGFPQPTGVTGPKHARGAQQPPGATGTFPCAAGRVGPPGSGNGPPGPPGPPS 923
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 541 GKEXG 544
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 924 GKDXG 927
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
collagen alpha 1(II) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A41182; A44885
R:Metaearanta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, A:
A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: A41182
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1419 <MET>
A:Cross-references: GB:M65161
R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartil
A:Reference number: A44885; MUID:91347939; PMID:1879363

Query Match 71.3%; Score 2189; DB 2; Length 1487;
Best Local Similarity 71.7%; Pred. No. 2.1e-98;
Matches 390; Conservative 39; Mismatches 115; Indels 0; Gaps 0;

QY	1	GSEPEGVRCBPGPPGAGAAAGPAGDPCADGCEPGCAKAGADGAPGIAGAPGFPGARGSPGPE	60
DB	384	GPEGAQSRGECGNPGSGFPAGASGNFTDGI PGAKGSAGAPGIAGAPGFPGRPPGPQ	443
QY	61	GPGEPPGKDGSGSPGAPGSGKDGTAKEBPQVGVGVEGPPGAPAGEGKPGARBPPTGILP	120
DB	444	GATGPLPKQAGSPGIAGFGDQCPKGETGPAGPQAGPAGPAGBEGKRGARPGCAGPI	503
QY	121	GPGERGGSRGPPGADGVAGPKGAPAGERSPGAPGKSGPGEAGRPCGKGLT	180
DB	504	GPPEERGAPGNRGFPQDGLAGPKGAPGERPGLAGPKGANGDFGRPGEGPLPGARGLT	563
QY	181	GSPGSPGDGKTGPPGAGEGRGPPGPPGARGGAGVMGFPKGAAGEPGKAGRGVP	240
DB	564	GRPCDAGPQKVGPSGAGEDGRGPPGPPGARGGAGVMGFPKGAAGEPGKAGEKGLA	623
QY	241	GPPGAVGPKDGBAGAEPPGPPGAPGAGERGEGPAGSPGPEGLPGPAGPPGAEAGKPGEE	300
DB	624	GAPLRLGPKDGTGAAGPPGSPGAPGERGEGQAGPSPGFOGLPGPPPPGEGGQGDQ	683
QY	301	GVPCDLGAPGSPGARGEPGFPGERGVGPPGAPGPPGADGAPDGDGAKGDAGAPGSGE	360
DB	684	GIPGEAGAPGLVGRBGRGFPGERGSPGAQGLQGPRGLPGTPTDGPKGAAGDPGPPGAQ	743
QY	361	GAPGLEMPGERGAAGLPFGPKGDRDAGPKGADGSPGKDGVRGLTGPIGPPPGAPAGDK	420
DB	744	GPPGLQMPGERGAAGIAGPKGDRGVGKPEGAPGKDGGRGLTGPIGPPPGAPAGKEX	803
QY	421	GESGSPGAPGTGARGAPCDGRGPPGPPGAPGPPGADGCEPGAKGEPDAGAKGDAGPP	480
DB	804	GEVFPSPGSGTGARGAPGEPGFTGTPPGPAGPAGPPGADGQPGKAGDQGEAGQKGDAGAP	863
QY	481	GPAGPAGPPGPIGVDPAGPAGKAGSAGPPGATGPPGAAGRVGPPGSPGDAGPPGPPGPA	540
DB	864	GPQPSGAPGPGGTGTGTGPKARGAQPPGATGPPGAGRVGPPGANGNPPGAPGPPGA	923
QY	541	GKEG 544	
DB	924	GKDG 927	

RESULT 8
CGRTIS
collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jul-1981 #sequence revision 13-Jul-1981 #text change 31-Mar-2000
C:Accession: A90559; A90552; A92023; A90353; A90566; A90357; A90362; A90379; A91209; A91102; A91103; A91104; A91105; A91106; A91107; A91108; A91109; A91110; A91111; A91112; A91113; A91114; A91115; A91116; A91117; A91118; A91119; A91120; A91121; A91122; A91123; A91124; A91125; A91126; A91127; A91128; A91129; A91130; A91131; A91132; A91133; A91134; A91135; A91136; A91137; A91138; A91139; A91140; A91141; A91142; A91143; A91144; A91145; A91146; A91147; A91148; A91149; A91150; A91151; A91152; A91153; A91154; A91155; A91156; A91157; A91158; A91159; A91160; A91161; A91162; A91163; A91164; A91165; A91166; A91167; A91168; A91169; A91170; A91171; A91172; A91173; A91174; A91175; A91176; A91177; A91178; A91179; A91180; A91181; A91182; A91183; A91184; A91185; A91186; A91187; A91188; A91189; A91190; A91191; A91192; A91193; A91194; A91195; A91196; A91197; A91198; A91199; A91200; A91201; A91202; A91203; A91204; A91205; A91206; A91207; A91208; A91209; A91210; A91211; A91212; A91213; A91214; A91215; A91216; A91217; A91218; A91219; A91220; A91221; A91222; A91223; A91224; A91225; A91226; A91227; A91228; A91229; A91230; A91231; A91232; A91233; A91234; A91235; A91236; A91237; A91238; A91239; A91240; A91241; A91242; A91243; A91244; A91245; A91246; A91247; A91248; A91249; A91250; A91251; A91252; A91253; A91254; A91255; A91256; A91257; A91258; A91259; A91260; A91261; A91262; A91263; A91264; A91265; A91266; A91267; A91268; A91269; A91270; A91271; A91272; A91273; A91274; A91275; A91276; A91277; A91278; A91279; A91280; A91281; A91282; A91283; A91284; A91285; A91286; A91287; A91288; A91289; A91290; A91291; A91292; A91293; A91294; A91295; A91296; A91297; A91298; A91299; A91300; A91301; A91302; A91303; A91304; A91305; A91306; A91307; A91308; A91309; A91310; A91311; A91312; A91313; A91314; A91315; A91316; A91317; A91318; A91319; A91320; A91321; A91322; A91323; A91324; A91325; A91326; A91327; A91328; A91329; A91330; A91331; A91332; A91333; A91334; A91335; A91336; A91337; A91338; A91339; A91340; A91341; A91342; A91343; A91344; A91345; A91346; A91347; A91348; A91349; A91350; A91351; A91352; A91353; A91354; A91355; A91356; A91357; A91358; A91359; A91360; A91361; A91362; A91363; A91364; A91365; A91366; A91367; A91368; A91369; A91370; A91371; A91372; A91373; A91374; A91375; A91376; A91377; A91378; A91379; A91380; A91381; A91382; A91383; A91384; A91385; A91386; A91387; A91388; A91389; A91390; A91391; A91392; A91393; A91394; A91395; A91396; A91397; A91398; A91399; A91400; A91401; A91402; A91403; A91404; A91405; A91406; A91407; A91408; A91409; A91410; A91411; A91412; A91413; A91414; A91415; A91416; A91417; A91418; A91419; A91420; A91421; A91422; A91423; A91424; A91425; A91426; A91427; A91428; A91429; A91430; A91431; A91432; A91433; A91434; A91435; A91436; A91437; A91438; A91439; A91440; A91441; A91442; A91443; A91444; A91445; A91446; A91447; A91448; A91449; A91450; A91451; A91452; A91453; A91454; A91455; A91456; A91457; A91458; A91459; A91460; A91461; A91462; A91463; A91464; A91465; A91466; A91467; A91468; A91469; A91470; A91471; A91472; A91473; A91474; A91475; A91476; A914

J. Biol. Chem. 242, 2572-2574, 1967
A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.
A;Reference number: A92029; MUID:67165368; PMID:4290711
A;Contents: CNBr2
A;Accession: A92029
A;Molecule type: protein
A;Residues: 20-55 <B02>
A;Experimental source: skin and tendon
R;Butler, W.T.; Ponds, S.L.
Biochemistry 10, 2076-2081, 1971
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino acid sequence of the hydroxylated alpha 1(I) chain of rat skin collagen.
A;Reference number: A90353; MUID:71263178; PMID:4327399
A;Contents: CNBr4
A;Accession: A90353
A;Molecule type: protein
A;Residues: 56-102 <B01>
A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The complete amino acid sequence of the hydroxylated alpha 1(I) chain of rat skin collagen.
A;Reference number: A90566; MUID:70085124; PMID:5411206
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <B02>
A;Experimental source: skin
R;Ballian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylated alpha 1(I) chain of rat skin collagen.
A;Reference number: A90357; MUID:72136131; PMID:4335087
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <B01>
A;Experimental source: skin
R;Ballian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylated alpha 1(I) chain of rat skin collagen.
A;Reference number: A90362; MUID:73006942; PMID:4342027
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <B02>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2946-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino acid sequence of the hydroxylated alpha 1(I) chain of rat skin collagen.
A;Reference number: A90379; MUID:74271984; PMID:4366532
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein
A;Residues: 419-567 <B03>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neutral salt-extracted rat skin collagen.
A;Reference number: A91209; MUID:74011954; PMID:4126850
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <S01>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain of rat skin collagen.
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <S02>
A;Experimental source: skin

A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequence of the alpha 1(I) chain of rat skin collagen.
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052 of the alpha 1(I) chain of rat skin collagen.
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C) are subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin collagen is: 1-9, 10-19, 20-55, 56-102, 103-139, 140-238, 239-418, 419-567, 568-651.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C;Keywords: blocked amino end, coiled coil, extracellular matrix, glycoprotein, hydroxylated, modified site, blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status F;9/Modified site: allysine (Lys) #status experimental
F;103,424,547/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match	Score	DB 1	Length	671					
Best Local Similarity	77.3%	Pred. No. 5.5e-97							
Matches	381	Conservative	28	Mismatches	36	Indels	48	Gaps	2
QY	1	GSEGEVGRGPPGPPGAGAGDPCAGDGEPCAGAGDAGAPGAGPFGGAGPFGP	60						
DB	200	GSEGEVGRGPPGPPGAGAGDPCAGDGEPCAGAGDAGAPGAGPFGGAGPFGP	259						
QY	61	GPGGPPGKDSGSPGAPSGKDTCAKGEPCPVGVEGPPGAGBEKPGARGEPTGLP	120						
DB	260	GPGGPPGKDSGSPGAPSGKDTCAKGEPCPVGVEGPPGAGBEKPGARGEPTGLP	319						
QY	121	GPGERGGPSRGPFGADGAVAGPKPAGERGSPGAPGKSPGAGRPGEAGLPGAKGLT	180						
DB	320	GPGERGGPSRGPFGADGAVAGPKPAGERGSPGAPGKSPGAGRPGEAGLPGAKGLT	379						
QY	181	GSPGSPGDKTGPFGAGDGRFPGPAGGAGVGMGPPGPKAAAGEPCKAGERGVP	240						
DB	380	GSPGSPGDKTGPFGAGDGRFPGPAGGAGVGMGPPGPKAAAGEPCKAGERGVP	439						
QY	241	GPGAVGPDGAGAGAGGPPGAPGAGGEGGPPGAGGEGGPPGAGGEGGPPGAGG	300						
DB	440	GPGAVGPDGAGAGAGGPPGAPGAGGEGGPPGAGGEGGPPGAGGEGGPPGAGG	499						
QY	301	GVPDGLGAPGSGAGSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	360						
DB	500	GVPDGLGAPGSGAGSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	559						
QY	361	GAPGLEGMPPGAGAGLPGPKDGRDAGPKGADGSPGKDGVRGLTGTTPGPPGAPG	420						
DB	560	GAPGLEGMPPGAGAGLPGPKDGRDAGPKGADGSPGKDGVRGLTGTTPGPPGAPG	583						
QY	421	GESGSPGAPGAGAGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	480						
DB	584	GESGSPGAPGAGAGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	631						
QY	481	GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	493						
DB	632	GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	644						

RESULT 9
B40333
collagen alpha 1(I) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: B40333
J. Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis embryogenesis.
A;Reference number: A40333; MUID:92011898; PMID:1918153
A;Accession: B40333
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1486 <SUA>
A;Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C;Keywords: coiled coil, extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <WVC>

	Qy	121	GPPGGRGFSGRGGPGADCVAGPVGKPAAGERGSPCPAGPKSPGEAAGRPGEAGLPGAKGLT	181
	Dd	513	GVVGRGAFGNRRGPFPSDGLFPFKGAQQRGVPVGSNGPKSQSDPRPCBPGLPGARGLT	572
	Qy	181	GSFSPGPDGKTGPPCPAGEDRGPDPGPCGARGEAQVMGFPGPKGAAGSPGKAGRGVGP	240
	Dd	573	GNPVQGPEGKLPLFCAPGEHORFPFGSITGIKQPWTMGLPGPKSNGDGPKFGAENP	632
	Qy	241	GPPGAVGAGKDXGEAAGPPGPGPAPAGERGEGPAGSPGFEGLPGPAGPBGAGKPGBE	300
	Dd	633	GVPGQRGAPGDKGVPTGPPGPPCLRGERGEQEPPTGTGQHHPGPPGPGSGGPGDGQ	692
	Qy	301	GVPGDLGAPGSGARGEGPFGERGVEGPPCPAGPPCADGAPGDDGAKDAGAPGPGSE	360
	Dd	693	GVPGGFAVGLPGPGERGNFGEEGEPGITGLPEKGMAGHGHDGPKSGSPGSGTPGDT	752
	Qy	361	GAPLEGMPGERGAAGLPCKPDGRDAGPKGADGSPGKDVGRLTGIPGPPGAPAGPGDK	420
	Dd	753	GPFGQLQMPGERGIAGTGPCKDRGGIGEKAECTAGNDGAGGLPGPLGPPGAGLLGEX	812
	Qy	421	GSBSGSPGAPPTGARCAPGDRGEPGPPGPAAGFAGPPGADGEPGAKGSPGDAGKDGAPP	480
	Dd	813	GBPGPRGLVPPSGRSNPGSRGNGPTCAVGFAGPQSDGQPGVKGPGSPGQKGDAOSP	872
	Qy	481	GPAGPAGPPGPIGDVCGAPGARGSGSAGPGCATGFCPGAAGRVCPPGSGDAGPPGPPA	540
	Dd	873	GPQGLAGSPGPHGPNVPLKGKRGTOGPPATGFGPSARGVGPPGAPAGPGPAGLGE	932
	Qy	541	GKEG 544	
	Dd	933	GKEG 936	
		RESULT 13		
		A43291	collagen alpha 2(I) chain precursor - mouse	
		C;Species:	Mus musculus (house mouse)	
		C;Date:	10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004	
		C;Accession:	A43291; A54328	
		R;Phillips,	C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.	
		Genomics	13, 1345-1346, 1992	
		A;Title:	Sequence analysis of a full-length cDNA for the murine pro alpha 2(I)	
		A;Reference	number: A43291; MUID:92372043; PMID:1505972	
		A;Accession:	A43291	
		A>Status:	preliminary; not compared with conceptual translation	
		A:Molecule	type: mRNA	
		A;Residues:	1-1373 <PHI>	
		A;Cross-references:	UNIPROT:Q01149; GB:X58251; NID:G50488; PIDN:CAA41205.1; PDB	
		A>Note:	sequence extracted from NCBI backbone (NCBI:P:112027)	
		R;Phillips,	C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.	
		J. invest.	Dermatol. 97, 980-984, 1991	
		A;Title:	Construction of a full-length murine Proalpha2(I) collagen cDNA by th	
		A;Reference	number: A54328; MUID:92084969; PMID:1748823	
		A;Accession:	A54328	
		A>Status:	preliminary; not compared with conceptual translation	
		A:Molecule	type: mRNA	
		A;Residues:	1-110 <PH2>	
		C;Genetics:		
		A;Gene:	COL1A2	
		C;Superfamily:	collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal	
		C;Keywords:	coiled coil; extracellular matrix; glycoprotein; trimer; triple hel	
		F;1145-1373/Domain:	fibrillar collagen carboxyl-terminal homology <FCC>	
		Query Match	62.4%; Score 1915; DB 1; Length 1373;	
		: Best Local Similarity	64.2%; Pred. No. 2.7e-85;	
		Matches 349; Conservative	45; Mismatches 150; Indels 0; Gaps 0;	
	Qy	1	GSEBPGVTGEGPPGPPAGAAGPADGCADCEPCAAGKADGACGTAGAGPFGARGSPGPE	60
	Dd	280	GPAGPAGPRGVEGLPGSLGSPGVGPPNPQNTGLTKAKGATLPGVAGAHGUPGPRGIPGA	339
	Qy	61	GPCCGPPGPKDGSBGPAPGSKGDTGAKGEPGVGVEGPPGAPAGEGPKGARGPFGTGLP	120


```

Db      340  GAAGATGATGVLGEPGAGSGKSGNKGESVGAQPPGSPGSEEGKRGSPGEAGSAGPA 399
QY      121  GPPGERGGSGRGGPPGADGVAGPKGPAERGSPPGAPKGSPPGAEAGLPGAKGLT 180
Db      400  GPFGLRGSFGRGLPGADGRAGVWGPNGKSGTGPAGIRPNFGDAGRPPGLMGRGLP 459
QY      181  GSPGSGPDGKTGTPPGPAGBEDGRPPGPPGARGAGVMGFPKPKGAAGSPGKAGERGVP 240
Db      460  GSPGNVGSKEGPGVGLPGTDRPGTIGPAGPRGEAGNIGFPKPKGSPGSDPGKPGRGHP 519
QY      241  GPPGAVCPAGKDEAGAGGPPGAGPAGERGSPAGSGPQFGLPGPAGPPGEGAGKPGEE 300
Db      520  GLAGARGAPGDNNGAQQGPPGQVQGGKEGQAPGPPGFQGLPGSPGTGTEVGPGR 579
QY      301  GVPEDLGAAPSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
Db      580  GLPGEFGLPGAPRGERTGTPGESGAAGSPGPTGSRGSPGAPDPDGNKGEAGVAGP 639
QY      361  GAPLEGMPGEGGAALPGPKGDRGADGPKGADGSPKDGVRGLTGP1GPPGPPAGAGDK 420
Db      640  GASGPGGLPGERGAAGTPGKGKGTGLRGDTGNTGRDARGIPGAVGAPGPPAGSGDR 699
QY      421  GESGSPGAGTARGANPDGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
Db      700  GEAGAAGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 759
QY      481  GPAGPAGPPGPIGDVGAAPGAKGARGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
Db      760  GPTGSGVAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 819
QY      541  GKEG 544
Db      820  GKEG 823

```

RESULT 14

I49607
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
 R:Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
 Dev. Dyn. 195, 113-120, 1992
 A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
 A:Reference number: I49607; PMID:93214071; PMID:1297453
 A:Accession: I49607
 A:Status: preliminary; translated from GB/EMBL/DDBT
 A:Molecule type: mRNA
 A:Residues: 1-1497 <RES>
 A:Cross-references: UNIPROT:Q61431; GB:L02918; NID:g309180; PIDN:AAA37440.1; PID:g309181
 C:Genetics:
 A:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 F:39-98/Domain: von Willebrand factor type C repeat homology <WVC>
 F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.2%; Score 1910; DB 2; Length 1497;
 Best Local Similarity 63.8%; Pred. No. 5e-85;
 Matches 347; Conservative 39; Mismatches 158; Indels 0; Gaps 0;

```

QY      1  GSEGPVGVRGEPGPPGAGAGPAGDPGADGPGKAGAGAGPAGPAGPAGPAGPAGP 60
Db      394  GPEGQQRGETGTPPGPAGSQGLPLGAVGTDGTPGRKATGSAAGTSGPPGLAGPPGSPGQ 453
QY      61  GPGGPPGPKDSEPCAPSKGDTGAKGPPGVGPPGPPGPPGPPGPPGPPGPPGPPGPP 120
Db      454  GSTGPGIRQSQDGPVPGFKGAGPKGPPGPHGIOGPPGPEEGKRPGRDPGVGPP 513
QY      121  GPPGERGSGRFPFGADGVAGPKGPPAGERGSPGPPGPPGPPGPPGPPGPPGPPGPP 180
Db      514  GPMGERGAPGNRFPFGSDGLPGPKQAQGERGPPVSSGPKGGQDPPRPPGPPGPPG 573

```

```

QY      181  GSPGSGPDGKTGTPPGPAGBEDGRPPGPPGARGAGVMGFPKPKGAAGSPGKAGERGVP 240
Db      574  GNPVGQEGEGKLGPLGAFGEDGRPPGSGIRGQPGSGMVGDPKSGSSGDLKPGEGAGNA 633
QY      241  GPPGAVCPAGKDEAGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
Db      634  GVPGRGAPGKDEGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 693
QY      301  GVPEDLGAAPSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
Db      694  GVPGPPGAVPLGPPKRGKNGPGERGEPGITGLPGKGMAGHGPPDPKNGNPPTTIGDT 753
QY      361  GAPLEGMPGEGGAALPGPKGDRGADGPKGADGSPKDGVRGLTGP1GPPGPPAGAGDK 420
Db      754  GPPGLQGMFGERGIAGTTPGKDRGIGKEKGAEGTAGNDGARGLPGPLPGPAGLLGAP 813
QY      421  GESGSPGAGTARGANPDGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
Db      814  GEPGPRGLVGPSPGSRKNGPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 873
QY      481  GPAGPAGPPGPIGDVGAAPGAKGARGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
Db      874  GFQGLAGSGFGPHGPHGVPLKGRGTQGPATGPPGATGPPGATGPPGATGPPGATGPP 933
QY      541  GKEG 544
Db      934  GKEG 937

```

RESULT 15

S59856
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
 C:Accession: S59856; S62120; S16373
 R:Toman, P.D.; de Crombrughe, B.
 Gene 147, 161-168, 1994
 A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA s
 A:Reference number: S59856; MUID:95011609; PMID:7926795
 A:Accession: S59856
 A:Molecule type: DNA
 A:Residues: 1-1464 <TOM>
 A:Cross-references: UNIPROT:P08121; EMBL:X52046
 R:Toman, D.
 submitted to the EMBL Data Library, November 1994
 A:Reference number: S62120
 A:Molecule type: DNA
 A:Residues: 1-866, G, 868-1464 <TOA>
 A:Cross-references: EMBL:X52046; NID:g575321; PIDN:CAA36279.1; PID:g575322
 R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
 Biochim. Biophys. Acta 1089, 241-243, 1991
 A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
 A:Reference number: S16176; MUID:91274355; PMID:2054384
 A:Accession: S16373
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1442-1464 <MET>
 A:Cross-references: EMBL:X57983; NID:g50476; PIDN:CAA41048.1; PID:g50477
 C:Genetics:
 A:Introns: 29/1, 95/3, 112/3, 150/3, 175/3, 193/3, 211/3, 229/3, 247/3, 265/3, 283/3, 298
 58/3, 673/3, 706/3, 742/3, 760/3, 778/3, 796/3, 814/3, 850/3, 868/3, 886/3, 940/3, 976/3;
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-154/Domain: propeptide #status predicted <PRO>
 F:155-1464/Domain: von Willebrand factor type C repeat homology <WVC>
 F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.1%; Score 1908; DB 2; Length 1464;
 Best Local Similarity 63.3%; Pred. No. 6.1e-85;
 Matches 346; Conservative 37; Mismatches 156; Indels 6; Gaps 2;

THIS PAGE BLANK (USPTO)

1	2966	96.6	1069	2	Q61AN8	homo sapien
2	2966	96.6	1461	2	Q76045	homo sapien
3	2966	96.6	1461	1	CALL_HUMAN	P076045
4	2966	96.6	1464	2	Q8N473	homo sapien
5	2940	95.8	1460	1	CALL_CANPA	Q8N473
6	2872	93.6	1453	2	Q63079	canis fami
7	2857	93.1	1453	1	CALL_MOUSE	Q63079
8	2857	93.1	1453	2	Q810J9	rattus norv
9	2775	90.4	1453	1	CALL_CHICK	mus musculu
10	2645.5	86.2	1425	2	Q6PCL3	mus musculu
11	2528	82.3	1450	2	Q9YIB4	gallus gall
12	2525	82.2	1445	2	Q93251	mus musculu
13	2525	82.2	1449	2	Q802B5	canis fami
14	2524	82.2	1449	2	Q640B2	canis fami
15	2484.5	80.9	1447	2	Q91B91	homo sapien
16	2307	75.1	1447	2	Q6U1J5	homo sapien
17	2298	74.9	1447	2	Q6P4U1	brachydanio
18	2231	72.7	1449	2	Q910C0	brachydanio
19	2225	72.5	1487	2	Q77753	oncorhynch
20	2224	72.4	1160	2	Q14046	canis fami
21	2224	72.4	1487	2	Q14047	homo sapien
22	2220	72.3	1418	2	Q28396	equus caball
23	2210	72.0	1449	2	Q6PE19	brachydanio
24	2208	71.9	1449	2	Q6NZ15	brachydanio
25	2202	71.7	1418	1	CALL2_HUMAN	homo sapien
26	2195	71.5	1269	2	Q7R2Z7	gallus gall
27	2192	71.4	1419	2	Q63123	rattus norv
28	2192	71.4	1419	2	Q80VY3	mus musculu
29	2192	71.4	1419	2	Q80X38	mus musculu
30	2192	71.4	1442	2	Q62031	mus musculu
31	2192	71.4	1442	2	Q62033	mus musculu

RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13 (1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379 (1991).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korcko J., Ala-Kokko L., De Paape A., Nuytink L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:
RT identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110 (1998).
RN [7]
RP SEQUENCE FROM N.A.
RX Korcko J.M., Earley J.J., Nuytink L., DePaape A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AB94054.2; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; VNC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS50184; VWF_C_2; 1.
DR Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30A78E21 CRC64;

DR	InterPro; IPR000895; Fib.collagen_C.
DR	InterPro; IPR009041; FMP_SGCI.
DR	InterPro; IPR001007; VWF_C.
DR	Pfam; PF01410; COLFI; 1.
DR	Pfam; PF01391; Collagen; 18.
DR	ProDom; PD000007; Clg_helix; 3.
DR	ProDom; PD002078; Fib.collagen_C; 1.
DR	SMART; SM00038; COLFI_1.
DR	SMART; SM00214; VWC; 1.
DR	PROSITE; PS01208; WFEC_1; 1.
DR	PROSITE; PS50184; WVEC_2; 1.
KW	Collagen.
SW	SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;
	Query Match 96.6%; Score 2966; DB 2; Length 1461;
	Best Local Similarity 95.6%; Pred. No. 1.le-106;
	Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
Qy	1 GSEGEVGRGEPPGPAGAAAGPDCADGEPKAGKADGAPGTAGAGPGPGARGSPGE 60
Dd	359 GSEGFQGVKRGPPGGPAGAAPGNFCADQPCKAKCANGAPGIAGNPFPGARGSGQG 418
Qy	61 GPGGPPGPKGSGERPACPSKDTCAKGEPGVVEGPPGPAEGEKPGARGEPPGTGLP 120
Dd	419 GPGGPPGPKGNSGEPACPGSKGDTCAKGEPGVVGVPQGPAGEGKRGARGBPGTGLP 478
Qy	121 GPPGERGGPGSRGPPGCADVAGPKPAGERGSPCPAGPKSPCEAGRCPGKAGLPGAAGLT 180
Dd	479 GFPGERGGPGSRGPPGDADVAGPKPAGERGSPCPAGPKSPCEAGRCPGKAGLPGAAGLT 538
Qy	181 GSPGSPGPDGKTGPPGAPGEDRPPGPPGARGEAGVMGMFGPKGAAGEPGKAGERGVP 240
Dd	539 GSPGSPGPDGKTGPPGAGQDGRPPGPPPCARGAQAGVMGMFGPKGAAGEPGKAGERGVP 598
Qy	241 GPPCAAGVPAKXDGAGAEAGPPGPAGPAGERGSGPAGSCFFGLPGAPGPPGEACKPGE 300
Dd	599 GPPCAAGVPAKXDGAGAEAGPPGPAGPAGERGSGPAGSCFFGLPGAPGPPGEACKPGE 658
Qy	301 GVPGDLGAPGPSARGBPFPGERGVEGPPGPAGPFCADGAPDDCGAKGDAGAPAGPGE 360

[illegible]

Db 659 GVPGLGAPGSGARGERGFPGRCVQPPGPGAGPRGANGAPNDGAKGDAGAPCSQ 718
Qy 361 GAPGLEHGPBRGAALPPGPKDGDAGPKGADGSPGKDGVRGLTGTPGPPGAPAGCDK 420
Db 719 GAPGLQGPBRGAALPPGPKDGDAGPKGADGSPGKDGVRGLTGTPGPPGAPAGCDK 778
Qy 421 GESGSPGAGPTGARGAGDGRGECPPGPGAGFPAGGAGDGPAGKGPAGDAGKGDAGPP 480
Db 779 GESGSPGAGPTGARGAGDGRGECPPGPGAGFPAGGAGDGPAGKGPAGDAGKGDAGPP 838
Qy 481 GPAGPAGPPTGIDVGAAPGAKGARGSPGAGTGPAGGAGRVGPPGSPGDAGPPGPGPA 540
Db 839 GPAGPAGPPTGIDVGAAPGAKGARGSPGAGTGPAGGAGRVGPPGSPGDAGPPGPGPA 898
Qy 541 KKEG 544
Db 899 KKEG 902
RESULT 3
CALL HUMAN
ID CALL HUMAN STANDARD; PRT: 1464 AA.
AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
AC Q15201; Q16050; Q7K230; Q7K234; Q8IVT5; Q9UML6; Q9UMW7;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Homo sapiens (Human),
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RP SEQUENCE FROM N.A., AND VARIANT SER-1434.
RA Daigleish R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE OF 1-589 FROM N.A.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RL Gene 67:105-115(1988).
RN (3)
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN (4)
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN (5)
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
RC TISSUE-Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN (6)
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;

RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
RL of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN (7)
RP SEQUENCE OF 472-607 FROM N.A.
RX PubMed=2981843;
RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
RT "Multixon deletion in an osteogenesis imperfecta variant with
RT increased type III collagen mRNA.";
RL J. Biol. Chem. 260:691-694(1985).
RN (8)
RP SEQUENCE OF 488-625 FROM N.A.
RX PubMed=3857621;
RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
RT I collagen chain in a lethal form of osteogenesis imperfecta.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
RN (9)
RP SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-II HIS-1277; ARG-1388 AND
RX 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
RX MEDLINE=93352646; PubMed=8349697;
RA Chessler S.D., Wallis G.A., Byers P.H.;
RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
RT chain of type I collagen result in defective chain association and
RT produce lethal osteogenesis imperfecta.";
RL J. Biol. Chem. 268:18218-18225(1993).
RN (10)
RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=88124208; PubMed=3340531;
RA Maekelae J.K., Raasina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN (11)
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN (12)
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to transcriptional
RT control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN (13)
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN (14)
RP SEQUENCE OF 33-52.
RX PubMed=2318855;
RA Rao V.H., Hollister D.W.;
RT "In vivo and in vitro noncovalent association of excised alpha 1 (I)
RT amino-terminal propeptides with mutant pro alpha 2(I) collagen chains
RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
RT VII.";
RL J. Biol. Chem. 265:6312-6317(1990).
RN (15)
RP SEQUENCE OF 156-183 FROM N.A.
RX PubMed=2767050;
RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,

RA Bateman J.F.;
 RT "A base substitution in the exon of a collagen gene causes alternative
 RT splicing and generates a structurally abnormal polypeptide in a
 RT patient with Ehlers-Danlos syndrome type VII.";
 RL EMBO J. 8:1705-1710(1989).
 RN [16]
 RP SEQUENCE OF 175-187 AND 274-289.
 RX PubMed=2169412;
 RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
 RA Muller P.K.;
 RT "A critical crosslink region in human-bone-derived collagen type I.
 RT Specific cleavage site at residue Leu95.";
 RL Eur. J. Biochem. 192:153-159(1990).
 RN [17]
 RP SEQUENCE OF 263-268.
 RC TISSUE=Skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected vertebrate
 RT collagens. A possible role of the carbohydrate in fibril formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [18]
 RP SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND
 RP 1143-1162 FROM N.A.
 RX PubMed=2374517;
 RA Labhard M.E., Hollister D.W.;
 RT "Segmental amplification of the entire helical and telopeptide regions
 RT of the cDNA for human alpha 1 (I) collagen.";
 RL Matrix 10:124-130(1990).
 RN [19]
 RP SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
 RX MEDLINE=83064528; PubMed=6183642;
 RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
 RT "Cloning and characterization of five overlapping cDNAs specific for
 RT the human pro alpha 1(I) collagen chain.";
 RL Nucleic Acids Res. 10:5925-5934(1982).
 RN [20]
 RP SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
 RX PubMed=2339700;
 RA Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
 RT "Variable expression of osteogenesis imperfecta in a nuclear family is
 RT explained by somatic mosaicism for a lethal point mutation in the
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
 RL Am. J. Hum. Genet. 46:1034-1040(1990).
 RN [21]
 RP SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
 RX MEDLINE=95187161; PubMed=7881420;
 RA Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
 RA Brunelli P.C., Mottes M.;
 RT "Severe (type III) osteogenesis imperfecta due to glycine
 RT substitutions in the central domain of the collagen triple helix.";
 RL Hum. Mol. Genet. 3:2201-2206(1994).
 RN [22]
 RP SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.
 RX PubMed=3170557;
 RA Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andreasen P.,
 RA Carboneau H., Nicholls A.C., Pope F.M., Byers P.H.;
 RT "Substitution of cysteine for glycine within the carboxyl-terminal
 RT telopeptide of the alpha 1 chain of type I collagen produces mild
 RT osteogenesis imperfecta.";
 RL J. Biol. Chem. 263:14605-14607(1988).
 RN [23]
 RP SEQUENCE OF 1440-1464 FROM N.A.
 RX MEDLINE=90110490; PubMed=2295701;
 RA Willing M.C., Cohn D.H., Byers P.H.;
 RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I
 RT collagen predicts an elongated pro alpha 1(I) chain and results in
 RT osteogenesis imperfecta type I.";
 RL J. Clin. Invest. 85:282-290(1990).
 RN [24]
 RP SEQUENCE OF 1454-1464 FROM N.A.
 RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
 RA Maatta A., Bornstein P., Penttinen R.P.;

RT "Highly conserved sequences in the 3'-untranslated region of the
 RT COL1A1 gene bind cell-specific nuclear proteins.";

Query Match 96.6%; Score 2966; DB 1; Length 1464;
 Best Local Similarity 95.6%; Pred. No. 1,1e-106;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPGAGAGPAGDPGADGSPGAGKADGAPGAGPFGARGSPGPE 60
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 362 GSEGPQGVGRGPPGPPGAGAGPAGDPGADGSPGAGKADGAPGAGPFGARGSPGPE 421
 QY 61 GFGGPPGPKGDSGEPGAPGSGKDTGAKGEPGVGVGPPGPRGEGKKGARCEPPTGLP 120
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 422 GFGGPPGPKGDSGEPGAPGSGKDTGAKGEPGVGVGPPGPRGEGKKGARCEPPTGLP 481
 QY 121 GPPGRRGGSGRGGFPFAGDVAGPKGPPAGERSPGPAGPKGSPGEGAGRPGEAGLPCA 180
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 482 GPPGRRGGSGRGGFPFAGDVAGPKGPPAGERSPGPAGPKGSPGEGAGRPGEAGLPCA 541
 QY 181 GSPGSGPDGKTGPPGPPGAGEDRPPGPPGARGGAGVMPGPKGAAGEPKGAGERGVP 240
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 542 GSPGSGPDGKTGPPGPPGAGEDRPPGPPGARGGAGVMPGPKGAAGEPKGAGERGVP 601
 QY 241 GPPGAVGPPAGKGEAGRPPGPPGAGRGERGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 602 GPPGAVGPPAGKGEAGRPPGPPGAGRGERGPPGPPGPPGPPGPPGPPGPPGPPGPP 661
 QY 301 GVPDLGAPGSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 662 GVPDLGAPGSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 721
 QY 361 GAPGLEGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 420
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 722 GAPGLEGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 781
 QY 421 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 782 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 841
 QY 481 GPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 842 GPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 901
 QY 541 GKEG 544
 DB |||||
 902 GKEG 905

RESULT 4

QBN473 ID QBN473 PRELIMINARY; PRT; 1464 AA.
 AC QBN473;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Alpha 1 type I collagen, preproprotein.
 GN Name=COL1A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullan S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RP [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC036531; AAH36531.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Clg helix; 3.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 139010 MW; B0581F8D1C89DDE8 CRC64;

Query Match 96.6%; Score 2966; DB 2; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1.1e-106;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPGAGAGPAGDPCADGEPGAKGADGAPGIAGAGPFGAGPSPGE 60
DB 362 GSEGEVGRGPPGPGAGAGPAGDPCADGEPGAKGADGAPGIAGAGPFGAGPSPGE 421
QY 61 GPGGPPGKGSRGPGAGSGKDTCAKGPVGVEGPPGAGGKPGAGCEPPTGLP 120
DB 422 GPGGPPGKGSRGPGAGSGKDTCAKGPVGVEGPPGAGGKPGAGCEPPTGLP 481

QY 121 GPPGGRGSGRGGPFGAGVAGPKGPGAGERSGPGAGPKGSGPGEAGRPGEAGLPGAKGLT 180
DB 482 GPPGGRGSGRGGPFGAGVAGPKGPGAGERSGPGAGPKGSGPGEAGRPGEAGLPGAKGLT 541

QY 181 GSPGSGPDGKTGTPGPDAGDGRPPGPPGARGAGVWGPFGPKGAAGPCKAGRGVP 240
DB 542 GSPGSGPDGKTGTPGPDAGDGRPPGPPGARGAGVWGPFGPKGAAGPCKAGRGVP 601

QY 241 GPPGAVGAGKDGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 300
DB 602 GPPGAVGAGKDGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 661

QY 301 GVPGLGAGPSPGARGEPFGPGERGVEGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGP 360
DB 662 GVPGLGAGPSPGARGEPFGPGERGVEGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGP 721

QY 361 GAPGLEGMWPGGAGLPGPKGDRGDAGPKGADGSGKDGVRGLTGPTGPPGAGAGDK 420
DB 722 GAPGLEGMWPGGAGLPGPKGDRGDAGPKGADGSGKDGVRGLTGPTGPPGAGAGDK 781

QY 421 GSGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 480
DB 782 GSGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 841

QY 481 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 540
DB 842 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 901

QY 541 GKEG 544
DB 902 GKEG 905

RESULT 5
CALL CANFA STANDARD; PRT; 1460 AA.
AC Q9XSJ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
RC TISSUE=Skin;
RX MEDLINE=21023337; PubMed=1147834; DOI=10.1006/abbi.2000.2099;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA and identification of a
RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case of
RT canine osteogenesis imperfecta";
RL Arch. Biochem. Biophys. 384:37-46 (2000).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains
CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
CC (OI).
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 VWF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF153062; AAD34619.1; -;
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Clg helix; 2.
DR ProDom; PD000007; Clg helix; 18.
DR ProDom; PD002078; Fib collagen_C; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen; Disease mutation; Extracellular matrix; Glycoprotein;
KW Hydroxylation; Pyridinolone carboxylic acid; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 22 By similarity.
FT PROPEP 23 157 N-terminal propeptide.
FT CHAIN 158 1214 Collagen alpha 1(I) chain.
FT PROPEP 1215 1460 C-terminal propeptide.
FT DOMAIN 34 92 VWF_C.
FT DOMAIN 158 174 Nonhelical region (N-terminal).
FT DOMAIN 175 1188 Triple-helical region.
FT DOMAIN 1189 1214 Nonhelical region (C-terminal).
FT SITE 741 743 Cell attachment site (Potential).

AC P11087; Q60635;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN Name=Colla1; Synonyms=Colal;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RX MEDLINE=96033240; PubMed=8535610;
 RA Li S.W., Khillan J., Prockop D.J.;
 RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
 of type I procollagen.";
 RL Matrix Biol. 14:593-595(1995).
 RN [2]
 RP SEQUENCE OF 518-1128 FROM N.A.
 RX MEDLINE=86137403; PubMed=3841523; DOI=10.1016/0378-1119(95)90329-4;
 RA French B.T., Lee W.-H., Maul G.G.;
 RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I) collagen
 protein.";
 RL Gene 39:311-312(1985).
 RN [3]
 RP SEQUENCE OF 735-1130 FROM N.A.
 RX MEDLINE=83141374; PubMed=6298597;
 RA Monson J.M., Friedman J., McCarthy B.J.;
 RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
 evidence for a mouse B1 element within the gene.";
 RL Mol. Cell. Biol. 2:1362-1371(1982).
 RN [4]
 RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
 RX MEDLINE=83157109; PubMed=6219867;
 RA Monson J.M., McCarthy B.J.;
 RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
 evidence for insertions or deletions in gene coding sequences.";
 RL DNA 1:59-69(1981).
 RN [5]
 RP SEQUENCE OF 1442-1453 FROM N.A.
 RX MEDLINE=88124276; PubMed=3340560;
 RA Mooslehner K., Harbers K.;
 RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size of
 the 3'-untranslated region.";
 RL Nucleic Acids Res. 16:773-773(1988).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
 bones. In bones the fibrils are mineralized with calcium
 hydroxyapatite.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U08020; AAA88912.1; -
 CC EMBL; X15896; CAA33904.1; -
 CC EMBL; M14423; AAA37333.1; -
 CC EMBL; M17491; AAA37334.1; -
 CC EMBL; X06753; CAA29927.1; -
 CC EMBL; K03036; AAA37332.1; -
 CC EMBL; K03029; AAA37332.1; JOINED.
 CC EMBL; K03030; AAA37332.1; JOINED.

DR EMBL; K03031; AAA37332.1; JOINED.
 DR EMBL; K03032; AAA37332.1; JOINED.
 DR EMBL; K03033; AAA37332.1; JOINED.
 DR EMBL; K03034; AAA37332.1; JOINED.
 DR EMBL; K03035; AAA37332.1; JOINED.
 DR PIR; S57243; S21626.
 DR MGD; MGI-88467; Colial.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR00885; Fib_collagen_C.
 DR InterPro; IPR009041; PMP_SGCI.
 DR InterPro; IPR01007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g_helix; 3.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM0038; COLFI; 1.
 DR SMART; SM00214; WVC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS0184; VWF_C_2; 1.
 KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;
 Pyridone carboxylic acid; Repeat; Signal; Structural protein.
 FT SIGNAL 1 22 N-terminal propeptide.
 FT PROPEP 23 151 Collagen alpha 1(I) chain.
 FT CHAIN 152 1207 C-terminal propeptide.
 FT PROPEP 1208 1453 VWFC.
 FT DOMAIN 29 87 Nonhelical region (N-terminal).
 FT DOMAIN 152 167 Triple-helical region.
 FT DOMAIN 168 1181 Nonhelical region (C-terminal).
 FT DOMAIN 1182 1207 Pyridone carboxylic acid (By
 FT MOD_RES 152 152 similarity).
 FT MOD_RES 160 160 Allysine (By similarity).
 FT MOD_RES 254 254 5-hydroxylysine (By similarity).
 FT MOD_RES 1153 1153 3-hydroxyproline (By similarity).
 FT CARBOHYD 56 56 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
 FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
 FT SITE 734 736 Cell attachment site (Potential).
 FT SITE 1082 1084 Cell attachment site (Potential).
 FT CONFLICT 1450 1450 A -> V (in Ref. 5).
 SQ SEQUENCE 1453 AA; 137944 MW; 3B802E535DF81808 CRC64;
 Query Match 93.1%; Score 2857; DB 1; Length 1453;
 Best Local Similarity 91.4%; Pred. No. 1.6e-102;
 Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;
 QY 1 GSEGEVGRGEPGPPAGAGADGADGEPGAGKAGADGAGTATAGATGAGTGGTGLP 60
 DB 351 GSEGEVGRGEPGPPAGAGAGADGADGEPGAGKAGADGAGTATAGATGAGTGGTGLP 410
 QY 61 GPGGPPGKGSDEPGAPGSGKGTGAKGEPGVGVGPPGAGGEGKPGAGGEPGTGLP 120
 DB 411 GPGGPPGKGSDEPGAPGSGKGTGAKGEPGVGVGPPGAGGEGKPGAGGEPGTGLP 470
 QY 121 GPPGERGGSGRGGFGADGAVAGKPGAGERSGPPGAGKSGPGGAGRGEAGLPGAKGLT 180
 DB 471 GPPGERGGSGRGGFGADGAVAGKPGAGERSGPPGAGKSGPGGAGRGEAGLPGAKGLT 530
 QY 181 GPGSGPGDGTGTPGPPAGGDRGPPGPPGARGGAGVGMGPPGKAGGEPKAGRGVP 240
 DB 531 GPGSGPGDGTGTPGPPAGGDRGPPGPPGARGGAGVGMGPPGKAGGEPKAGRGVP 590
 QY 241 GPPGAVGPGKDGAGAGGPPGPPAGGERGEGPAGSGPFGGLPGGAPGGEAGKPGEE 300
 DB 591 GPPGAVGPGKDGAGAGGPPGPPAGGERGEGPAGSGPFGGLPGGAPGGEAGKPGEE 650
 QY 301 GVPGLGAPGSGARGEGPFGGERGVEGPPGPPGAGGADGAGDDGAKGDAGAPGPGSE 360
 DB 651 GVPGLGAPGSGARGEGPFGGERGVEGPPGPPGAGGADGAGDDGAKGDAGAPGPGSE 710
 QY 361 GAPGLEGMPGARGAAGLPGPKGDRGDAGPKGADSGPKGVRLTGTPTGPPGAPGADK 420


```
Db 711 GAPGLQMPGGERGAAGLPGKGRGDAGPKGADGSPKDGARGLTGPIGPPGAPAGDK 770
QY 421 GESGSPGAGFTGARGAGPDRGPPGPPGAGFAGPPGADGEPGAKGEPGAGAKGDAGPP 480
Db 771 GEAGSPGPPGFTGARGAGPDRGPPGPPGAGFAGPPGADGEPGAKGEPGAGTGVKGDAGPP 830
QY 481 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFPGAAGRVGPPSPGSDAGPPGPPGA 540
Db 831 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFPGAAGRVGPPSPGSDAGPPGPPGV 890
QY 541 GKEG 544
Db 891 GKEG 894

RESULT 8
Q810J9 PRELIMINARY; PRT; 1453 AA.
ID AC Q810J9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-MAR-2004 (TrEMBLrel. 24, Last sequence update)
DE Procollagen, type I, alpha 1.
GN Name=COL1A1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Basa S.A., McWane P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RC STRAIN=FVB/N; TISSUE=Colon;
RX STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050014; AAH50014.1; -.
DR MGD; MGI:88467; Colla1.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR009041; FWP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_1; 1.
```

```
DR PROSITE; PS0184; VWF_2; 1.
KW Collagen.
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;
Query Match 93.1%; Score 2857; DB 2; Length 1453;
Best Local Similarity 91.4%; Pred. No. 1.6e-102;
Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;
QY 1 GSEGGVGRGPPGPPGAGAGPAGDPCGADGEPGAKGADGAGPIAGAPCFPGAGPSGPE 60
Db 351 GSEGGVGRGPPGPPGAGAGPAGNPGADGQPGKANGAPGTAGAGFFPGARGPSGPQ 410
QY 61 GFSGPPGPKDGSSEGFSGKSGDTGAKGEPGVGVGEGPPGAGBEGKFGKAGPEPTGLP 120
Db 411 GFSGPPGPKDGSSEGFSGKSGDTGAKGEPGATGVQGGPPGAGBEGKFGKAGPEPTGLP 470
QY 121 GPPGERGPPGSRGPPGADGAVAGPKGPPAGERGSPGPPAGPKGSPGEGRGERGLPKAKGLT 180
Db 471 GPPGERGPPGSRGPPGADGAVAGPKGPPAGERGSPGPPAGPKGSPGEGRGERGLPKAKGLT 530
QY 181 GSPGSPGPDGKTGTPPGPAGEDRPPGPPGARGSEAGVMGPPGKGAAGEPKAGRGVVP 240
Db 531 GSPGSPGPDGKTGTPPGPAGEDRPPGPPGARGSEAGVMGPPGKGAAGEPKAGRGVVP 590
QY 241 GPPGAVGPAKDGAGAGGPPGPPGAGERGEEGPPGAGSPGFEGLFGPAGPFCGAGKPGEE 300
Db 591 GPPGAVGPAKDGAGAGGPPGPPGAGERGEEGPPGAGSPGFEGLFGPAGPFCGAGKPGEE 650
QY 301 GVPDGLGAPGSPGARGEPGPFGRGVEGPPGPPGADGAPDDGAKGDAGAPGAPGSE 360
Db 651 GVPDGLGAPGSPGARGEPGPFGRGVEGPPGPPGADGAPDDGAKGDAGAPGAPGSE 710
QY 361 GAPGLEGMPGRRGAAGLPGPKDGDAGPKGADGSPGKDVGRGLTGPITGPPGAPAGDK 420
Db 711 GAPGLQMPGGERGAAGLPGPKDGDAGPKGADGSPGKDVGRGLTGPITGPPGAPAGDK 770
QY 421 GESGSPGAGTARGAPGDRGPPGPPGAGFAGPPGADGEPGAKGEPGAGTGVKGDAGPP 480
Db 771 GEAGSPGPPGFTGARGAGPDRGPPGPPGAGFAGPPGADGEPGAKGEPGAGTGVKGDAGPP 830
QY 481 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFPGAAGRVGPPSPGSDAGPPGPPGA 540
Db 831 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFPGAAGRVGPPSPGSDAGPPGPPGV 890
QY 541 GKEG 544
Db 891 GKEG 894

RESULT 9
CALL_CHICK STANDARD; PRT; 1453 AA.
ID AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RC SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=88056316; PubMed=3678834; DOI=10.1016/0378-1119(87)90159-4;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RL of the chicken pro alpha 1(I) collagen mRNA."
RN [2]
RC SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
```


RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
 RT "Unusual DNA sequences located within the promoter region and the
 RT first intron of the chicken pro-alpha 1(I) collagen gene."; [3]
 RL J. Biol. Chem. 262:13323-13332(1987).
 RN [3]
 RP SEQUENCE OF 152-1187.
 RX MEDLINE=82231995; PubMed=7093229;
 RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H.,
 RA Gross J.;
 RA "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
 RT complete primary structure of the helical portion of the chick skin
 RT collagen alpha 1(I) chain."; [4]
 RL Biochemistry 21:2048-2055(1982).
 RN [4]
 RP SEQUENCE OF 1200-1205.
 RX MEDLINE=72243016; PubMed=5047697;
 RA Eyre D.R., Glimcher M.J.;
 RA "Evidence for a previously undetected sequence at the carboxyterminus
 RT of the alpha 1 chain of chicken bone collagen."; [5]
 RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
 RN [5]
 RP SEQUENCE OF 981-1453 FROM N.A.
 RX MEDLINE=81160715; PubMed=6927845;
 RA Fuller F., Boedtker H.;
 RA "Sequence determination and analysis of the 3' region of chicken pro-
 RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
 RT including the carboxy-terminal propeptide sequences."; [6]
 RL Biochemistry 20:996-1006(1981).
 RN [6]
 RP SEQUENCE OF 1311-1453 FROM N.A.
 RX MEDLINE=80134546; PubMed=6987088; DOI=10.1016/0014-5793(80)80761-7;
 RA Shewalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
 RA Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
 RA "Nucleotide sequence of a collagen cDNA-fragment coding for the
 RT carboxyl end of pro alpha 1(I)-chains."; [6]
 RL FEBS Lett. 111:61-65(1980).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 CC (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
 CC bones. In bones the fibrils are mineralized with calcium
 CC hydroxyapatite.
 CC -!- PTM: Proline residues at the third position of the tripeptide
 CC repeating unit (G-X-Y) are hydroxylated in some or all of the
 CC chains. Pro-1153 is the only 3-hydroxypro and the only
 CC hydroxylated proline in position X.
 CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -!- SIMILARITY: Contains 1 WFVC domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M17839; AAA48704.1; -;
 DR EMBL; M17838; AAA48704.1; JOINED.
 DR EMBL; V00401; CA23695.1; -;
 DR EMBL; M10571; AAA48671.1; ALT_SEQ.
 DR EMBL; M17607; AAA48672.1; -;
 DR PIR; A27179; A27179.
 DR PIR; I50629; I50629.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; WFVC.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; WVC; 1.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; WFVC.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; WVC; 1.
 DR ProDom; PD000007; Clg_helix; 2.
 DR ProDom; PD000207; Fib_collagen_C; 1.

DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; WVC; 1.
 DR PROSITE; PS01208; WFVC_1; 1.
 DR PROSITE; PS0184; WFVC_2; 1.
 KW Collagen; Direct protein sequencing; Extracellular matrix;
 KW Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;
 KW Signal; Structural protein.
 FT SIGNAL 1 22 N-terminal propeptide.
 FT PROPEP 23 151 Collagen alpha 1(I) chain.
 FT CHAIN 152 1205 C-terminal propeptide.
 FT PROPEP 1206 1453 WFVC.
 FT DOMAIN 31 89 Pyrrolidone carboxylic acid.
 FT MOD_RES 152 152 Allysine (By similarity).
 FT MOD_RES 160 160 5-hydroxylysine (By similarity).
 FT MOD_RES 254 254 5-hydroxylysine (Potential).
 FT MOD_RES 1081 1081 Hydroxyproline (Potential).
 FT MOD_RES 1097 1097 5-hydroxylysine (Potential).
 FT MOD_RES 1153 1153 3-hydroxyproline.
 FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
 FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
 FT CONFLICT 1187 1187 P -> L (in Ref. 5).
 FT CONFLICT 1441 1441 Q -> H (in Ref. 6).
 SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;
 Query Match 90.4%; Score 2775; DB 1; Length 1453;
 Best Local Similarity 89.0%; Pred. No. 2,3e-99;
 Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;
 QY 1 GSEGEVGRGEPGPPGAGADGADGECAGKAGDAGPAGPAGPARGPSGPE 60
 DB 351 GSEGEVGRGEPGPPGAGADGADGECAGKAGDAGPAGPARGPSGPE 60
 QY 61 GPGGPPGKDGSGEPGAGSGKGTGAKGEGPVGVEGPGPGAGEGKRGARGPGLP 120
 DB 411 GPGGPPGKDGSGEPGAGSGKGTGAKGEGPVGVEGPGPGAGEGKRGARGPGLP 120
 QY 121 GPPGEGGPGSRGPPGADGVAGPKGAGERSGPGAGPKSGPAGRGAGLPGAKGLT 180
 DB 471 GPGGEGGPGSRGPPGADGVAGPKGAGERSGPGAGPKSGPAGRGAGLPGAKGLT 180
 QY 181 GSPGSGPDGKTGTPGPPGAGEDGPPGPPGARGGAGVGMGPPGKGAAGEPKGAGRGVP 240
 DB 531 GSPGSGPDGKTGTPGPPGAGEDGPPGPPGARGGAGVGMGPPGKGAAGEPKGAGRGVP 240
 QY 241 GPPGAVGPPAGKDGAGAGGPPGPPGAGPAGERGEEGPPGSGPGLPGAGPKGEE 300
 DB 591 GPPGAVGPPAGKDGAGAGGPPGPPGAGPAGERGEEGPPGSGPGLPGAGPKGEE 300
 QY 301 GVPGLGAPGSGARGBPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
 DB 651 GVPGLGAPGSGARGBPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
 QY 361 GARGLEGMPGEGAGLPGPKGDRGADGPKGADGSPGKGVGLTGTPTGPPGPPGAGDK 420
 DB 711 GARGLEGMPGEGAGLPGPKGDRGADGPKGADGSPGKGVGLTGTPTGPPGPPGAGDK 420
 QY 421 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
 DB 771 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 830
 QY 481 GPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
 DB 831 GPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
 QY 541 GKEG 544
 DB 891 GKZG 894
 RESULT 10
 Q6PCL3 PRELIMINARY; PRT; 1225 AA.
 ID Q6PCL3

AC Q6PCL3;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Collal protein.
 GN Name=Collal;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RX MEDLINE=2238825; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RA Strausberg R.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC059281; AAH59281.1; -;
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008885; Fib_collagen_C.
 DR InterPro; IPR009041; PMP_SGCI.
 DR InterPro; IPR010007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 13.
 DR ProDom; PD000093; VWC; 1.
 DR ProDom; PD000007; Clg_helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS01208; VWF_C_2; 1.
 DR PROSITE; PS01208; VWF_C_3; 1.
 DR PROSITE; PS01208; VWF_C_4; 1.
 DR PROSITE; PS01208; VWF_C_5; 1.
 DR PROSITE; PS01208; VWF_C_6; 1.
 DR PROSITE; PS01208; VWF_C_7; 1.
 DR PROSITE; PS01208; VWF_C_8; 1.
 DR PROSITE; PS01208; VWF_C_9; 1.
 DR PROSITE; PS01208; VWF_C_10; 1.
 DR PROSITE; PS01208; VWF_C_11; 1.
 DR PROSITE; PS01208; VWF_C_12; 1.
 DR PROSITE; PS01208; VWF_C_13; 1.
 DR PROSITE; PS01208; VWF_C_14; 1.
 DR PROSITE; PS01208; VWF_C_15; 1.
 DR PROSITE; PS01208; VWF_C_16; 1.
 DR PROSITE; PS01208; VWF_C_17; 1.
 DR PROSITE; PS01208; VWF_C_18; 1.
 DR PROSITE; PS01208; VWF_C_19; 1.
 DR PROSITE; PS01208; VWF_C_20; 1.
 DR PROSITE; PS01208; VWF_C_21; 1.
 DR PROSITE; PS01208; VWF_C_22; 1.
 DR PROSITE; PS01208; VWF_C_23; 1.
 DR PROSITE; PS01208; VWF_C_24; 1.
 DR PROSITE; PS01208; VWF_C_25; 1.
 DR PROSITE; PS01208; VWF_C_26; 1.
 DR PROSITE; PS01208; VWF_C_27; 1.
 DR PROSITE; PS01208; VWF_C_28; 1.
 DR PROSITE; PS01208; VWF_C_29; 1.
 DR PROSITE; PS01208; VWF_C_30; 1.
 DR PROSITE; PS01208; VWF_C_31; 1.
 DR PROSITE; PS01208; VWF_C_32; 1.
 DR PROSITE; PS01208; VWF_C_33; 1.
 DR PROSITE; PS01208; VWF_C_34; 1.
 DR PROSITE; PS01208; VWF_C_35; 1.
 DR PROSITE; PS01208; VWF_C_36; 1.
 DR PROSITE; PS01208; VWF_C_37; 1.
 DR PROSITE; PS01208; VWF_C_38; 1.
 DR PROSITE; PS01208; VWF_C_39; 1.
 DR PROSITE; PS01208; VWF_C_40; 1.
 DR PROSITE; PS01208; VWF_C_41; 1.
 DR PROSITE; PS01208; VWF_C_42; 1.
 DR PROSITE; PS01208; VWF_C_43; 1.
 DR PROSITE; PS01208; VWF_C_44; 1.
 DR PROSITE; PS01208; VWF_C_45; 1.
 DR PROSITE; PS01208; VWF_C_46; 1.
 DR PROSITE; PS01208; VWF_C_47; 1.
 DR PROSITE; PS01208; VWF_C_48; 1.
 DR PROSITE; PS01208; VWF_C_49; 1.
 DR PROSITE; PS01208; VWF_C_50; 1.
 DR PROSITE; PS01208; VWF_C_51; 1.
 DR PROSITE; PS01208; VWF_C_52; 1.
 DR PROSITE; PS01208; VWF_C_53; 1.
 DR PROSITE; PS01208; VWF_C_54; 1.
 DR PROSITE; PS01208; VWF_C_55; 1.
 DR PROSITE; PS01208; VWF_C_56; 1.
 DR PROSITE; PS01208; VWF_C_57; 1.
 DR PROSITE; PS01208; VWF_C_58; 1.
 DR PROSITE; PS01208; VWF_C_59; 1.
 DR PROSITE; PS01208; VWF_C_60; 1.
 DR PROSITE; PS01208; VWF_C_61; 1.
 DR PROSITE; PS01208; VWF_C_62; 1.
 DR PROSITE; PS01208; VWF_C_63; 1.
 DR PROSITE; PS01208; VWF_C_64; 1.
 DR PROSITE; PS01208; VWF_C_65; 1.
 DR PROSITE; PS01208; VWF_C_66; 1.
 DR PROSITE; PS01208; VWF_C_67; 1.
 DR PROSITE; PS01208; VWF_C_68; 1.
 DR PROSITE; PS01208; VWF_C_69; 1.
 DR PROSITE; PS01208; VWF_C_70; 1.
 DR PROSITE; PS01208; VWF_C_71; 1.
 DR PROSITE; PS01208; VWF_C_72; 1.
 DR PROSITE; PS01208; VWF_C_73; 1.
 DR PROSITE; PS01208; VWF_C_74; 1.
 DR PROSITE; PS01208; VWF_C_75; 1.
 DR PROSITE; PS01208; VWF_C_76; 1.
 DR PROSITE; PS01208; VWF_C_77; 1.
 DR PROSITE; PS01208; VWF_C_78; 1.
 DR PROSITE; PS01208; VWF_C_79; 1.
 DR PROSITE; PS01208; VWF_C_80; 1.
 DR PROSITE; PS01208; VWF_C_81; 1.
 DR PROSITE; PS01208; VWF_C_82; 1.
 DR PROSITE; PS01208; VWF_C_83; 1.
 DR PROSITE; PS01208; VWF_C_84; 1.
 DR PROSITE; PS01208; VWF_C_85; 1.
 DR PROSITE; PS01208; VWF_C_86; 1.
 DR PROSITE; PS01208; VWF_C_87; 1.
 DR PROSITE; PS01208; VWF_C_88; 1.
 DR PROSITE; PS01208; VWF_C_89; 1.
 DR PROSITE; PS01208; VWF_C_90; 1.
 DR PROSITE; PS01208; VWF_C_91; 1.
 DR PROSITE; PS01208; VWF_C_92; 1.
 DR PROSITE; PS01208; VWF_C_93; 1.
 DR PROSITE; PS01208; VWF_C_94; 1.
 DR PROSITE; PS01208; VWF_C_95; 1.
 DR PROSITE; PS01208; VWF_C_96; 1.
 DR PROSITE; PS01208; VWF_C_97; 1.
 DR PROSITE; PS01208; VWF_C_98; 1.
 DR PROSITE; PS01208; VWF_C_99; 1.
 DR PROSITE; PS01208; VWF_C_100; 1.
 DR PROSITE; PS01208; VWF_C_101; 1.
 DR PROSITE; PS01208; VWF_C_102; 1.
 DR PROSITE; PS01208; VWF_C_103; 1.
 DR PROSITE; PS01208; VWF_C_104; 1.
 DR PROSITE; PS01208; VWF_C_105; 1.
 DR PROSITE; PS01208; VWF_C_106; 1.
 DR PROSITE; PS01208; VWF_C_107; 1.
 DR PROSITE; PS01208; VWF_C_108; 1.
 DR PROSITE; PS01208; VWF_C_109; 1.
 DR PROSITE; PS01208; VWF_C_110; 1.
 DR PROSITE; PS01208; VWF_C_111; 1.
 DR PROSITE; PS01208; VWF_C_112; 1.
 DR PROSITE; PS01208; VWF_C_113; 1.
 DR PROSITE; PS01208; VWF_C_114; 1.
 DR PROSITE; PS01208; VWF_C_115; 1.
 DR PROSITE; PS01208; VWF_C_116; 1.
 DR PROSITE; PS01208; VWF_C_117; 1.
 DR PROSITE; PS01208; VWF_C_118; 1.
 DR PROSITE; PS01208; VWF_C_119; 1.
 DR PROSITE; PS01208; VWF_C_120; 1.
 DR PROSITE; PS01208; VWF_C_121; 1.
 DR PROSITE; PS01208; VWF_C_122; 1.
 DR PROSITE; PS01208; VWF_C_123; 1.
 DR PROSITE; PS01208; VWF_C_124; 1.
 DR PROSITE; PS01208; VWF_C_125; 1.
 DR PROSITE; PS01208; VWF_C_126; 1.
 DR PROSITE; PS01208; VWF_C_127; 1.
 DR PROSITE; PS01208; VWF_C_128; 1.
 DR PROSITE; PS01208; VWF_C_129; 1.
 DR PROSITE; PS01208; VWF_C_130; 1.
 DR PROSITE; PS01208; VWF_C_131; 1.
 DR PROSITE; PS01208; VWF_C_132; 1.
 DR PROSITE; PS01208; VWF_C_133; 1.
 DR PROSITE; PS01208; VWF_C_134; 1.
 DR PROSITE; PS01208; VWF_C_135; 1.
 DR PROSITE; PS01208; VWF_C_136; 1.
 DR PROSITE; PS01208; VWF_C_137; 1.
 DR PROSITE; PS01208; VWF_C_138; 1.
 DR PROSITE; PS01208; VWF_C_139; 1.
 DR PROSITE; PS01208; VWF_C_140; 1.
 DR PROSITE; PS01208; VWF_C_141; 1.
 DR PROSITE; PS01208; VWF_C_142; 1.
 DR PROSITE; PS01208; VWF_C_143; 1.
 DR PROSITE; PS01208; VWF_C_144; 1.
 DR PROSITE; PS01208; VWF_C_145; 1.
 DR PROSITE; PS01208; VWF_C_146; 1.
 DR PROSITE; PS01208; VWF_C_147; 1.
 DR PROSITE; PS01208; VWF_C_148; 1.
 DR PROSITE; PS01208; VWF_C_149; 1.
 DR PROSITE; PS01208; VWF_C_150; 1.
 DR PROSITE; PS01208; VWF_C_151; 1.
 DR PROSITE; PS01208; VWF_C_152; 1.
 DR PROSITE; PS01208; VWF_C_153; 1.
 DR PROSITE; PS01208; VWF_C_154; 1.
 DR PROSITE; PS01208; VWF_C_155; 1.
 DR PROSITE; PS01208; VWF_C_156; 1.
 DR PROSITE; PS01208; VWF_C_157; 1.
 DR PROSITE; PS01208; VWF_C_158; 1.
 DR PROSITE; PS01208; VWF_C_159; 1.
 DR PROSITE; PS01208; VWF_C_160; 1.
 DR PROSITE; PS01208; VWF_C_161; 1.
 DR PROSITE; PS01208; VWF_C_162; 1.
 DR PROSITE; PS01208; VWF_C_163; 1.
 DR PROSITE; PS01208; VWF_C_164; 1.
 DR PROSITE; PS01208; VWF_C_165; 1.
 DR PROSITE; PS01208; VWF_C_166; 1.
 DR PROSITE; PS01208; VWF_C_167; 1.
 DR PROSITE; PS01208; VWF_C_168; 1.
 DR PROSITE; PS01208; VWF_C_169; 1.
 DR PROSITE; PS01208; VWF_C_170; 1.
 DR PROSITE; PS01208; VWF_C_171; 1.
 DR PROSITE; PS01208; VWF_C_172; 1.
 DR PROSITE; PS01208; VWF_C_173; 1.
 DR PROSITE; PS01208; VWF_C_174; 1.
 DR PROSITE; PS01208; VWF_C_175; 1.
 DR PROSITE; PS01208; VWF_C_176; 1.
 DR PROSITE; PS01208; VWF_C_177; 1.
 DR PROSITE; PS01208; VWF_C_178; 1.
 DR PROSITE; PS01208; VWF_C_179; 1.
 DR PROSITE; PS01208; VWF_C_180; 1.
 DR PROSITE; PS01208; VWF_C_181; 1.
 DR PROSITE; PS01208; VWF_C_182; 1.
 DR PROSITE; PS01208; VWF_C_183; 1.
 DR PROSITE; PS01208; VWF_C_184; 1.
 DR PROSITE; PS01208; VWF_C_185; 1.
 DR PROSITE; PS01208; VWF_C_186; 1.
 DR PROSITE; PS01208; VWF_C_187; 1.
 DR PROSITE; PS01208; VWF_C_188; 1.
 DR PROSITE; PS01208; VWF_C_189; 1.
 DR PROSITE; PS01208; VWF_C_190; 1.
 DR PROSITE; PS01208; VWF_C_191; 1.
 DR PROSITE; PS01208; VWF_C_192; 1.
 DR PROSITE; PS01208; VWF_C_193; 1.
 DR PROSITE; PS01208; VWF_C_194; 1.
 DR PROSITE; PS01208; VWF_C_195; 1.
 DR PROSITE; PS01208; VWF_C_196; 1.
 DR PROSITE; PS01208; VWF_C_197; 1.
 DR PROSITE; PS01208; VWF_C_198; 1.
 DR PROSITE; PS01208; VWF_C_199; 1.
 DR PROSITE; PS01208; VWF_C_200; 1.
 DR PROSITE; PS01208; VWF_C_201; 1.
 DR PROSITE; PS01208; VWF_C_202; 1.
 DR PROSITE; PS01208; VWF_C_203; 1.
 DR PROSITE; PS01208; VWF_C_204; 1.
 DR PROSITE; PS01208; VWF_C_205; 1.
 DR PROSITE; PS01208; VWF_C_206; 1.
 DR PROSITE; PS01208; VWF_C_207; 1.
 DR PROSITE; PS01208; VWF_C_208; 1.
 DR PROSITE; PS01208; VWF_C_209; 1.
 DR PROSITE; PS01208; VWF_C_210; 1.
 DR PROSITE; PS01208; VWF_C_211; 1.
 DR PROSITE; PS01208; VWF_C_212; 1.
 DR PROSITE; PS01208; VWF_C_213; 1.
 DR PROSITE; PS01208; VWF_C_214; 1.
 DR PROSITE; PS01208; VWF_C_215; 1.
 DR PROSITE; PS01208; VWF_C_216; 1.
 DR PROSITE; PS01208; VWF_C_217; 1.
 DR PROSITE; PS01208; VWF_C_218; 1.
 DR PROSITE; PS01208; VWF_C_219; 1.
 DR PROSITE; PS01208; VWF_C_220; 1.
 DR PROSITE; PS01208; VWF_C_221; 1.
 DR PROSITE; PS01208; VWF_C_222; 1.
 DR PROSITE; PS01208; VWF_C_223; 1.
 DR PROSITE; PS01208; VWF_C_224; 1.
 DR PROSITE; PS01208; VWF_C_225; 1.
 DR PROSITE; PS01208; VWF_C_226; 1.
 DR PROSITE; PS01208; VWF_C_227; 1.
 DR PROSITE; PS01208; VWF_C_228; 1.
 DR PROSITE; PS01208; VWF_C_229; 1.
 DR PROSITE; PS01208; VWF_C_230; 1.
 DR PROSITE; PS01208; VWF_C_231; 1.
 DR PROSITE; PS01208; VWF_C_232; 1.
 DR PROSITE; PS01208; VWF_C_233; 1.
 DR PROSITE; PS01208; VWF_C_234; 1.
 DR PROSITE; PS01208; VWF_C_235; 1.
 DR PROSITE; PS01208; VWF_C_236; 1.
 DR PROSITE; PS01208; VWF_C_237; 1.
 DR PROSITE; PS01208; VWF_C_238; 1.
 DR PROSITE; PS01208; VWF_C_239; 1.
 DR PROSITE; PS01208; VWF_C_240; 1.
 DR PROSITE; PS01208; VWF_C_241; 1.
 DR PROSITE; PS01208; VWF_C_242; 1.
 DR PROSITE; PS01208; VWF_C_243; 1.
 DR PROSITE; PS01208; VWF_C_244; 1.
 DR PROSITE; PS01208; VWF_C_245; 1.
 DR PROSITE; PS01208; VWF_C_246; 1.
 DR PROSITE; PS01208; VWF_C_247; 1.
 DR PROSITE; PS01208; VWF_C_248; 1.
 DR PROSITE; PS01208; VWF_C_249; 1.
 DR PROSITE; PS01208; VWF_C_250; 1.
 DR PROSITE; PS01208; VWF_C_251; 1.
 DR PROSITE; PS01208; VWF_C_252; 1.
 DR PROSITE; PS01208; VWF_C_253; 1.
 DR PROSITE; PS01208; VWF_C_254; 1.
 DR PROSITE; PS01208; VWF_C_255; 1.
 DR PROSITE; PS01208; VWF_C_256; 1.
 DR PROSITE; PS01208; VWF_C_257; 1.
 DR PROSITE; PS01208; VWF_C_258; 1.
 DR PROSITE; PS01208; VWF_C_259; 1.
 DR PROSITE; PS01208; VWF_C_260; 1.
 DR PROSITE; PS01208; VWF_C_261; 1.
 DR PROSITE; PS01208; VWF_C_262; 1.
 DR PROSITE; PS01208; VWF_C_263; 1.
 DR PROSITE; PS01208; VWF_C_264; 1.
 DR PROSITE; PS01208; VWF_C_265; 1.
 DR PROSITE; PS01208; VWF_C_266; 1.
 DR PROSITE; PS01208; VWF_C_267; 1.
 DR PROSITE; PS01208; VWF_C_268; 1.
 DR PROSITE; PS01208; VWF_C_269; 1.
 DR PROSITE; PS01208; VWF_C_270; 1.
 DR PROSITE; PS01208; VWF_C_271; 1.
 DR PROSITE; PS01208; VWF_C_272; 1.
 DR PROSITE; PS01208; VWF_C_273; 1.
 DR PROSITE; PS01208; VWF_C_274; 1.
 DR PROSITE; PS01208; VWF_C_275; 1.
 DR PROSITE; PS01208; VWF_C_276; 1.
 DR PROSITE; PS01208; VWF_C_277; 1.
 DR PROSITE; PS01208; VWF_C_278; 1.
 DR PROSITE; PS01208; VWF_C_279; 1.
 DR PROSITE; PS01208; VWF_C_280; 1.
 DR PROSITE; PS01208; VWF_C_281; 1.
 DR PROSITE; PS01208; VWF_C_282; 1.
 DR PROSITE; PS01208; VWF_C_283; 1.
 DR PROSITE; PS01208; VWF_C_284; 1.
 DR PROSITE; PS01208; VWF_C_285; 1.
 DR PROSITE; PS01208; VWF_C_286; 1.
 DR PROSITE; PS01208; VWF_C_287; 1.
 DR PROSITE; PS01208; VWF_C_288; 1.
 DR PROSITE; PS01208; VWF_C_289; 1.
 DR PROSITE; PS01208; VWF_C_290; 1.
 DR PROSITE; PS01208; VWF_C_291; 1.
 DR PROSITE; PS01208; VWF_C_292; 1.
 DR PROSITE; PS01208; VWF_C_293; 1.
 DR PROSITE; PS01208; VWF_C_294; 1.
 DR PROSITE; PS01208; VWF_C_295; 1.
 DR PROSITE; PS01208; VWF_C_296; 1.
 DR PROSITE; PS01208; VWF_C_297; 1.
 DR PROSITE; PS01208; VWF_C_298; 1.
 DR PROSITE; PS01208; VWF_C_299; 1.
 DR PROSITE; PS01208; VWF_C_300; 1.
 DR PROSITE; PS01208; VWF_C_301; 1.
 DR PROSITE; PS01208; VWF_C_302; 1.
 DR PROSITE; PS01208; VWF_C_303; 1.
 DR PROSITE; PS01208; VWF_C_304; 1.
 DR PROSITE; PS01208; VWF_C_305; 1.
 DR PROSITE; PS01208; VWF_C_306; 1.
 DR PROSITE; PS01208; VWF_C_307; 1.
 DR PROSITE; PS01208; VWF_C_308; 1.
 DR PROSITE; PS01208; VWF_C_309; 1.
 DR PROSITE; PS01208; VWF_C_310; 1.
 DR PROSITE; PS01208; VWF_C_311; 1.
 DR PROSITE; PS01208; VWF_C_312; 1.
 DR PROSITE; PS01208; VWF_C_313; 1.
 DR PROSITE; PS01208; VWF_C_314; 1.
 DR PROSITE; PS01208; VWF_C_315; 1.
 DR PROSITE; PS01208; VWF_C_316; 1.
 DR PROSITE; PS01208; VWF_C_317; 1.
 DR PROSITE; PS01208; VWF_C_318; 1.
 DR PROSITE; PS01208; VWF_C_319; 1.
 DR PROSITE; PS01208; VWF_C_320; 1.
 DR PROSITE; PS01208; VWF_C_321; 1.
 DR PROSITE; PS01208; VWF_C_322; 1.
 DR PROSITE; PS01208; VWF_C_323; 1.
 DR PROSITE; PS01208; VWF_C_324; 1.
 DR PROSITE; PS01208; VWF_C_325; 1.
 DR PROSITE; PS01208; VWF_C_326; 1.
 DR PROSITE; PS01208; VWF_C_327; 1.
 DR PROSITE; PS01208; VWF_C_328; 1.
 DR PROSITE; PS01208; VWF_C_329; 1.
 DR PROSITE; PS01208; VWF_C_330; 1.
 DR PROSITE; PS01208; VWF_C_331; 1.
 DR PROSITE; PS01208; VWF_C_332; 1.
 DR PROSITE; PS01208; VWF_C_333; 1.
 DR PROSITE; PS01208; VWF_C_334; 1.
 DR PROSITE; PS01208; VWF_C_335; 1.
 DR PROSITE; PS01208; VWF_C_336; 1.
 DR PROSITE; PS01208; VWF_C_337; 1.
 DR PROSITE; PS01208; VWF_C_338; 1.
 DR PROSITE; PS01208; VWF_C_339; 1.
 DR PROSITE; PS01208; VWF_C_340; 1.
 DR PROSITE; PS01208; VWF_C_341; 1.
 DR PROSITE; PS01208; VWF_C_342; 1.
 DR PROSITE; PS01208; VWF_C_343; 1.
 DR PROSITE; PS01208; VWF_C_344; 1.
 DR PROSITE; PS01208; VWF_C_345; 1.
 DR PROSITE; PS01208; VWF_C_346; 1.
 DR PROSITE; PS01208; VWF_C_347; 1.
 DR PROSITE; PS01208; VWF_C_348; 1.
 DR PROSITE; PS01208; VWF_C_349; 1.
 DR PROSITE; PS01208; VWF_C_350; 1.
 DR PROSITE; PS01208; VWF_C_351; 1.
 DR PROSITE; PS01208; VWF_C_352; 1.
 DR PROSITE; PS01208; VWF_C_353; 1.
 DR PROSITE; PS01208; VWF_C_354; 1.
 DR PROSITE; PS01208; VWF_C_355; 1.
 DR PROSITE; PS01208; VWF_C_356; 1.
 DR PROSITE; PS01208; VWF_C_357; 1.
 DR PROSITE; PS01208; VWF_C_358; 1.
 DR PROSITE; PS01208; VWF_C_359; 1.
 DR PROSITE; PS01208; VWF_C_360; 1.
 DR PROSITE; PS01208; VWF_C_361; 1.
 DR PROSITE; PS01208; VWF_C_362; 1.
 DR PROSITE; PS01208; VWF_C_363; 1.
 DR PROSITE; PS01208; VWF_C_364; 1.
 DR PROSITE; PS01208; VWF_C_365; 1.
 DR PROSITE; PS01208; VWF_C_366; 1.
 DR PROSITE; PS01208; VWF_C_367; 1.
 DR PROSITE; PS01208; VWF_C_368; 1.
 DR PROSITE; PS01208; VWF_C_369; 1.
 DR PROSITE; PS01208; VWF_C_370; 1.
 DR PROSITE; PS01208; VWF_C_371; 1.
 DR PROSITE; PS01208; VWF_C_372; 1.
 DR PROSITE; PS01208; VWF_C_373; 1.
 DR PROSITE; PS01208; VWF_C_374; 1.
 DR PROSITE; PS01208; VWF_C_375; 1.
 DR PROSITE; PS01208; VWF_C_376; 1.
 DR PROSITE; PS01208; VWF_C_377; 1.
 DR PROSITE; PS01208; VWF_C_378; 1.
 DR PROSITE; PS01208; VWF_C_379; 1.
 DR PROSITE; PS01208; VWF_C_380; 1.
 DR PROSITE; PS01208; VWF_C_381; 1.
 DR PROSITE; PS01208; VWF_C_382; 1.
 DR PROSITE; PS01208; VWF_C_383; 1.
 DR PROSITE; PS01208; VWF_C_384; 1.
 DR PROSITE; PS01208; VWF_C_385; 1.
 DR PROSITE; PS01208; VWF_C_386; 1.
 DR PROSITE; PS01208; VWF_C_387; 1.
 DR PROSITE; PS01208; VWF_C_388; 1.
 DR PROSITE; PS01208; VWF_C_389; 1.
 DR PROSITE; PS01208; VWF_C_390; 1.
 DR PROSITE; PS01208; VWF_C_391; 1.
 DR PROSITE; PS01208; VWF_C_392; 1.
 DR PROSITE; PS01208; VWF_C_393; 1.
 DR PROSITE; PS01208; VWF_C_394; 1.
 DR PROSITE; PS01208; VWF_C_395; 1.
 DR PROSITE; PS01208; VWF_C_396; 1.
 DR PROSITE; PS01208; VWF_C_397; 1.
 DR PROSITE; PS01208; VWF_C_398; 1.
 DR PROSITE; PS01208; VWF_C_399; 1.
 DR PROSITE; PS01208; VWF_C_400; 1.
 DR PROSITE; PS01208; VWF_C_401; 1.
 DR PROSITE; PS01208; VWF_C_402; 1.
 DR PROSITE; PS01208; VWF_C_403; 1.
 DR PROSITE; PS01208; VWF_C_404; 1.
 DR PROSITE; PS01208; VWF_C_405; 1.
 DR PROSITE; PS01208; VWF_C_406; 1.
 DR PROSITE; PS01208; VWF_C_407; 1.
 DR PROSITE; PS01208; VWF_C_408; 1.
 DR PROSITE; PS01208; VWF_C_409; 1.
 DR PROSITE; PS01208; VWF_C_410; 1.
 DR PROSITE; PS01208; VWF_C_411; 1.
 DR PROSITE; PS01208; VWF_C_412; 1.
 DR PROSITE; PS01208; VWF_C_413; 1.
 DR PROSITE; PS01208; VWF_C_414; 1.
 DR PROSITE; PS01208; VWF_C_415; 1.
 DR PROSITE; PS01208; VWF_C_416; 1.
 DR PROSITE; PS01208; VWF_C_417; 1.
 DR PROSITE; PS01208; VWF_C_418; 1.
 DR PROSITE; PS01208; VWF_C_419; 1.
 DR PROSITE; PS01208; VWF_C_420; 1.
 DR PROSITE; PS01208; VWF_C_421; 1.
 DR PROSITE; PS01208; VWF_C_422; 1.
 DR PROSITE; PS01208; VWF_C_423; 1.
 DR PROSITE; PS01208; VWF_C_424; 1.
 DR PROSITE; PS01208; VWF_C_425; 1.
 DR PROSITE; PS01208; VWF_C_426; 1.

KW Collagen. 1450 AA; 137563 MW; ABP8A74841B87B7C CRC64;
SQ SEQUENCE 1450 AA; 137563 MW; ABP8A74841B87B7C CRC64;
Query Match 82.3%; Score 2528; DB 2; Length 1450;
Best Local Similarity 80.3%; Pred. No. 6.6e-90;
Matches 437; Conservative 43; Mismatches 64; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPGAGAGPAGDPCADGEPGAGKAGDAGAGTGGAGPFGGAGPSPGE 60
DB 348 GSEGEVGRGPPGPPGAGAGPAGDPCADGEPGAGKAGDAGAGTGGAGPFGGAGPSPGE 407
QY 61 GPGGPPGKDSGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 120
DB 408 GPAGAPGKNNNGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 467
QY 121 GPPGGRGGSGRGGPAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 180
DB 468 GPAGGGRGGSGRGGPAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 527
QY 181 GSPGSGPDGKTGPPGAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 240
DB 528 GSPGSGPDGKTGPPGAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 587
QY 241 GPGGAVGAGKDSGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 300
DB 588 GPGGAVGAGKDSGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 647
QY 301 GVPGLGAGPAGSGAGGEPGAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 360
DB 648 GAPDAGGPGSGRGGPAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 707
QY 361 GAPGLGMPGSRGAGLPGKPGDGDAGPKGADGSGKDGVRGLTGTGPGPAGAGPAGDK 420
DB 708 GPPGLGMPGSRGAGLPGKPGDGDAGPKGADGSGKDGVRGLTGTGPGPAGAGPAGDK 767
QY 421 GESGSPGAGTGAGGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 480
DB 768 GEGGSPGAGTGAGGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 827
QY 481 GPAGPAGPPTGVDGAGPAGKAGSGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 540
DB 828 GPAGPAGPPTGVDGAGPAGKAGSGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 587
QY 541 GKEG 544
DB 888 GKEG 891
RESULT 12
O93251 PRELIMINARY; PRT; 1445 AA.
AC O93251;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
GN Name=alpha 1 type I collagen;
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99294154; PubMed=10367734; DOI=10.1016/S0945-053X(99)00005-0;
RX Asahina K., Uchida R., Ohara M., Yoshizato K.;
RT "Cell-type specific and thyroid hormone-dependent expression of genes
of alpha 1 and alpha 2(I) collagen in intestine during
amphibian metamorphosis.";
RL Matrix Biol. 18:89-103(1999).
DR EMBL; AB015440; BAA29028.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cyclopasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR01007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_1; 1.
DR PROSITE; PS0184; VWF_2; 1.
KW Collagen.
SQ SEQUENCE 1445 AA; 137251 MW; F59BB550C9873F04 CRC64;
Query Match 82.2%; Score 2525; DB 2; Length 1445;
Best Local Similarity 81.1%; Pred. No. 8.6e-90;
Matches 441; Conservative 37; Mismatches 66; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPGAGAGPAGDPCADGEPGAGKAGDAGAGTGGAGPFGGAGPSPGE 60
DB 347 GPDGPGGAGGPPGAGGAGPAGGAGPAGGAGPAGGAGPAGGAGPAGGAGPAGGAGPAGG 406
QY 61 GPGGPPGKDSGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 120
DB 407 GPGGPPGKNNNGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 466
QY 121 GPPGGRGGSGRGGPAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 180
DB 467 GPAGGGRGGSGRGGPAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 526
QY 181 GSPGSGPDGKTGPPGAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 240
DB 527 GSPGSGPDGKTGPPGAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 586
QY 241 GPGGAVGAGKDSGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 300
DB 587 GPGGAVGAGKDSGEPGAGKSGDTCAGKEPVGVEGPPGAGGEGKPGAGGEPGPTGLP 646
QY 301 GVPGLGAGPAGSGAGGEPGAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 360
DB 647 GAPDVGSPGSGRGGPAGDVGAGPAGKPGAGGSGPAGKPGSGPAGGAGRPGGAGLPGAKGLT 706
QY 361 GAPGLGMPGSRGAGLPGKPGDGDAGPKGADGSGKDGVRGLTGTGPGPAGAGPAGDK 420
DB 707 GPSGLGMPGSRGAGLPGKPGDGDAGPKGADGSGKDGVRGLTGTGPGPAGAGPAGDK 766
QY 421 GESGSPGAGTGAGGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 480
DB 767 GEAGPAGPAGTGAGGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 826
QY 481 GPAGPAGPPTGVDGAGPAGKAGSGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 540
DB 827 GAAGPAGPAGPAGGAGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGPAGGSGP 886
QY 541 GKEG 544
DB 887 GKEG 890
RESULT 13
Q802B5 PRELIMINARY; PRT; 1449 AA.
AC Q802B5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collal-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

OC Xenopodinae; Xenopus.
 RN [1] NCBI_TaxID=8355;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 initiative.";
 RL Dev. Dyn. 225:384-391 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX Klein S., Strausberg R.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC049829; AAH49829.1;
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005737; C:cvtolasm; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR00885; Fib_collagen_C.
 DR InterPro; IPR009041; PMP_SGCI.
 DR InterPro; IPR001007; VWFC_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; VWC; 1.
 DR ProDom; PD000007; Clg_helix; 4.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWFC_1; UNKNOWN_1.
 DR PROSITE; PS0184; VWFC_2; 1.
 KW Collagen.
 SQ SEQUENCE 1449 AA; 137464 MW; 4B32E20BECCE978C CRC64;
 Query Match 82.2%; Score 2525; DB 2; Length 1449;
 Best Local Similarity 80.3%; Pred.No. 8.6e-90;
 Matches 437; Conservative 50; Mismatches 57; Indels 0; Gaps 0;
 QY 1 GSEGGVGEPPGPPAGAGPAGDPGADGPGFAGKAGADGAPGAGPFGPAGPSPGPE 60
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 60
 QY 348 GSDGPGGGRGEPGAPGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 407
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 407
 QY 61 GPGGPPGKDSGEPGAPGKSGDTGAKGEPGVPVGGPPGAGGKPGAGGPPGTLPL 120
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 120
 QY 408 GPGGPPGKGNNGEPGAPGAGKAGGEPGAGVQVGGPFGPFGSGRGRGSGRGPAGPP 467
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 467

QY 121 GPPGERGGPSRGFPAGDVGAGPKEPAGERGSPGAPGPKSGPCEAGRPGCEAGLPKAGLIT 180
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 180
 QY 468 GPAGERGGPSRGFPAGDVGAGPKEPAGERGSPGAPGPKSGPCEAGRPGCEAGLPKAGLIT 527
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 527
 QY 181 GSPGSGPPDGKTPGPPGAGEDGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 240
 QY 528 GSPGSGPSGKTPGPPGAGEDGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 587
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 587
 QY 241 GPPGAVGPPAGKDEAGAEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 300
 QY 588 GPPGAVGLPGKDGADAGAQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 647
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 647
 QY 301 GYPGDLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 360
 QY 648 GYPGDPVPSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 707
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 707
 QY 361 GAPGLEGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 420
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 420
 QY 708 GPPGLQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 767
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 767
 QY 421 GSPGSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 480
 QY 768 GEAGPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 827
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 827
 QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 540
 QY 828 GPAGTAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 887
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 887
 QY 541 GKEG 544
 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 891
 QY 888 GKEG 891

RESULT 14

Q640B2
 ID Q640B2 PRELIMINARY; PRT; 1449 AA.
 AC Q640B2;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8354;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Whole body;
 RX PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.

RC	TISSUE=whole body;	Query Match	82.2%;	Score 2524;	DB 2;	Length 1449;
RA	Klein S., Gernhard D.S.;	Best Local Similarity	80.3%;	Pred. No. 9.4e-90;		
RL	Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.	Matches 437;	Conservative 48;	Mismatches 59;	Indels 0;	Gaps 0;
DR	EMBL; BC082718; AAH82718.1; -					
KW	Hypothetical protein.					
SQ	SEQUENCE 1449 AA; 137539 MW; 45663981687A3B4B CRC64;					
QY	1	GSEGEVGRGEPGPPGAGAAAGADGADCEPGAKGADGAPGTAGAGPFGFCARGPSPE	60			
DB	348	PDGPGQGRGEPGAPGQAGAGPSGNSDQAGAKGATGAGTAGAGPFGFCARGAPGQ	407			
QY	61	PGPGPPGPKGDSGSPGAPGSGKDTGAKGEPGVGVEGPPGAGBEGKPGARGEPGFTGLP	120			
DB	408	PGPGAPGPKGNNGEPFGAQCNGKEPGAKGESGAGVQGGPPGSGEGRKSGRGEPPGAPGP	467			
QY	121	GPPEGRGGSRGFPFGADGVAGPKGAPGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLT	180			
DB	468	GPAGERGAPGSRGFPFGSDGASGPKGGPGERGVPVGPAGANGSPGSGRPGEPLPGAKGLT	527			
QY	181	GSPGSPGPDGKTGTPGPGAGEDGRPGPPGPGARGEAGVMGPPGPKGAAGEPKGAGERVPP	240			
DB	528	GSPGSPGPDGKTGAGAPGQDGRPGPPGPGARGQSGVMGPPGPKGAAGEPKSGERGVA	587			
QY	241	GPPGAVGPAKDKGAGABGPPGPGAPGAGERGEPGAGSPGEGULPGPAGPPGEAKPGEE	300			
DB	588	GPPGAVGLPGKDGAGAGQGGPPGAPGSGERGEQGPAGGFGQLPGPPGAPAGESCKPGEQ	647			
QY	301	GVPCDLGAPGSGARGEPGPPGGRGVGPPGPGADGAPDDGAKGDAGAPCAPGSE	360			
DB	648	GVPCDVFSGPAGARGERGFPGERGAQGPFGPQGPGRGNGAPGNDGAKGEAGACAPGQ	707			
QY	361	GAPGLEGMPSRGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGTPGPPGAPAGDK	420			
DB	708	GPPGLQGMPSRGAAGLPGAKGDRGQGVKSGDGAAPKDGVRGLTGTPGPPGAPAGDK	767			
QY	421	GESGFSPPAGTGTARGAPCDGRGEPGPPGAPGAPGADGEPGAKGEPGDAGAKGDAGPP	480			
DB	768	GEAGPAGPAGTGTARGAPGERSGPPGAPGAPGADGQFGAKGEGQDAGAKGDAGPP	827			
QY	481	GPAGPAGPPGPIGVGAPGAKGAGSAGPPGATGPPGAAGRVGPPGSGDAGPPGPPGPA	540			
DB	828	GPAGPTGAPGAPGALGATGPKARGAPGPPGATGFPGATGVRGPPGAPAGNSGPPGSPGA	887			
QY	541	GKEG 544				
DB	888	GKEG 891				

RESULT 15	
Q9IB91	
ID	Q9IB91
AC	PRELIMINARY;
DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT	01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE	Type I collagen alpha 1.
GN	Name=COL1A1;
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC	Xenopodinae; Xenopus.
OC	NCBI_TaxID=8355;
[1]	
RP	SEQUENCE FROM N.A.
RP	MEDLINE=20365486; PubMed=10910131;
RA	Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RA	"Expression and characterization of Xenopus type I collagen alpha 1
RT	(COL1A1) during embryonic development.;"
RT	Dev. Growth Differ. 42:249-256(2000).
RL	

DR	ENBL; AB034701; BAA94972.1; -;
DR	GO; GO:0005581; C:c collagen; IEA.
DR	GO; GO:0005737; C:cyttoplasm; IEA.
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR	GO; GO:0006817; P:phosphate transport; IEA.
DR	InterPro; IPR008161; clg_helix.
DR	InterPro; IPR008160; Collagen.
DR	InterPro; IPR000885; Fib_collagen_C.
DR	InterPro; IPR009041; PMP_SGCI.
DR	InterPro; IPR001007; VWF_C.
DR	Pfam; PF01410; COLFI; 1.
DR	Pfam; PF01391; Collagen; 18.
DR	Pfam; PF00093; WVC; 1.
DR	ProDom; PD000007; clg_helix; 5.
DR	ProDom; PD002078; Fib_collagen_C; 1.
DR	SMART; SM00038; COLFI; 1.
DR	SMART; SM00214; WVC; 1.
DR	PROSITE; PS01208; VWFC_1; UNKNOWN_1.
DR	PROSITE; PS50184; VWFC_2; 1.
KW	Collagen.
SQ	SEQUENCE 1447 AA; 137445 MW; A4A6DD2B4158B38B CRC64;
Query Match	
Best Local Similarity 79.9%; Score 2484.5; DB 2; Length 1447;	
Matches 433; Conservative 50; Mismatches 60; Indels 1; Gaps 1;	
Qy	1 GSEGPGEVGRGPPGGPAGAGADPGADGEPGAKGADGAPGIAGAPFGCARGPSGPE 60
Dd	348 GSDPPQGGRGEPAPCGAAGAASGNFSGSDQPQGAKGATGPIAGAPFPFCARGAPQAQ 407
Qy	61 GRGPPGPKDSCSEPAPGSKGDITGAKGEGPVGVGPPGAGREGKPGARGECPGTGLP 120
Dd	408 GPGSPGPKNGNBPAGQNKGEAGAKEPPAGVGQPPGSGEGKKSGKEP-PAGPP 466
Qy	121 GPCEERGPGSRGPPGADGVAGPGKPAGERGSPGAPGKSPGAGRPCEAGLPKAAGLT 180
Dd	467 GPAGERGGPSRGPFPGSDGASGPKGAPCERGVPVAGPKGSSGESGRFGEPCLPGAAGLT 526
Qy	181 GSPGSPGPDOKTTPPGPADGEDRPGPPGAPGARAGVMWGPPKGAAGEPKAAGERVVP 240
Dd	527 GSPGSPSGDKTGTPAGAPGQDGRAGPPGPPGARGQSGVMGPPKGAAGEPGKNGEKGA 586
Qy	241 GPCAVCPAGKDGAGAEAGPPGAPGAGEEAGSPGSPGEGLPGPAGPGCEACKPGEE 300
Dd	587 GPPGAVGLPKRDGDAGAQGPPGPPAGPAGERGEQGPPGFQGLPGSPGAGESKPGEQ 646
Qy	301 GVPCDLIAPGPSARGPPGPPGPERGVGPPPGPAPGPGADGAPDDGAKGDAGAPGASE 360
Dd	647 GVPCDVGPSGAPGARGERGPPGEGAQQPPGQCARGSNGAPNDGAKGEAGAAPGQG 706
Qy	361 GAPGLEWMPGERGAALPGPKDRDGDAGPKADGSPGKDGVRLGTGPTGPPGAPAGDGK 420
Dd	707 GSPGLQMGPERSSSLPGAKDGRDQGVKSGDGTGPKDGVRLGTGPTGPPGPGAPGDK 766
Qy	421 GESGPSGAPTGTARGAPGDRGEPGPPGAPGAGPPGADGPPGAKGEPGDAGAKGDAGPP 480
Dd	767 GEAGPAGPAGTGTARGAPGERGEPAPEPAPGAFGPPGADGQFKAAGEGDSGAKGDAGPS 826
Qy	481 GPAPGAPGPPGIDGVGAPGAKGARGSGAPPGCATGFPGAAGRVGPPGSDAGPPGPPGPA 540
Dd	827 GLPGTAPGAPGALGSPGPKGARGAPGPPCATGFPGAAGRVGPPGPNAGPPGSPGSPA 886
Qy	541 GKEX 544
Dd	887 GKEX 890

Search completed: June 17, 2005, 15:14:00
Job time : 82.6937 secs

THIS PAGE BLANK (USPTO)